Here Today, Gone Tomorrow? Toward an Understanding of Fade-out in Early Childhood Education Programs

John A. List*   Haruka Uchida‡

September 27, 2024

Abstract

An unsettling stylized fact is that decorated early childhood education programs improve cognitive skills in the short-term, but lose their efficacy after a few years. We implement a field experiment with two stages of randomization to explore the underpinnings of the fade-out effect. We first randomly assign preschool access to children, and then partner with the local school district to randomly assign the same children to classmates throughout elementary school. We find that the fade-out effect is critically-linked to the share of classroom peers assigned to preschool access—with enough treated peers the classic fade-out effect is muted. Our results highlight a paradoxical insight: while the fade-out effect has been viewed as a devastating critique of early childhood programs, our results highlight that fade-out is a key rational for providing early education to *all* children. This is because human capital accumulation is inherently a social activity, leading early education programs to deliver their largest benefits at scale when everyone receives such programs.

*University of Chicago, Australian National University, NBER, RFF, and IZA; jlist@uchicago.edu

‡University of Chicago; uchida@uchicago.edu. We thank Shruti Jha, Anya Samek, Adeline Sutton, and Kristin Troutman. We are grateful to Marco Castillo, Justin Holz, and Xianglong Kong for helpful comments and discussion. Yujin Lee and Francesca Pagnotta provided outstanding research assistance.

1 Introduction

At least since Mincer (1958), economists have recognized that investments in human capital can be analyzed similarly to investments in physical capital. Schultz (1961) expanded Mincer's original idea, arguing that human capital is a critical factor in improving productivity and economic growth. Becker (1962) took the baton from there, further developing human capital theory by exploring the economic implications of investing in education, training, and health. His work highlighted how such investments can lead to higher earnings and better job opportunities. This work, along with several other seminal contributions, laid the underpinnings for understanding how human capital contributes to economic growth and individual prosperity (see Heckman and Mosso, 2014; Attanasio et al., 2022).

As a whole, the corpus of work highlights that, overall, human capital is vital for both individual and societal progress, and that the foundations for human capital accumulation are laid early on. For example, recent work argues that investing in early childhood education (ECE) can yield significant returns (e.g. Yoshikawa et al., 2013; Elango et al., 2015). As the argument goes, since the early years—birth to five years old—are marked by significant brain development, children are highly receptive to learning and experiences, which shape their cognitive, emotional, and social development. In this manner, ECE is argued to provide a strong foundation for later academic success: children who attend quality early childhood programs are better prepared for school, exhibit improved literacy and numeracy skills, and are more likely to succeed academically. In addition, ECE can help to reduce the gap between socially advantaged and disadvantaged children, promoting equity and social justice.

While early childhood programs are often seen as an essential policy tool to enhance human capital and address social inequities, recently a consensus has emerged that paints a skeptical picture: program evaluations often find that initial boosts in cognitive skills fade out by the first few years of formal schooling.[1] Much of the literature has consequently attributed the value of early childhood interventions to its long-term impacts on "non-cognitive" skills, such as externalizing behavior (Heckman et al., 2013). Yet given that both cognitive and non-cognitive skills have been shown to play an important role in later-life outcomes such as labor market success (Heckman et al., 2006), it remains important to understand why cognitive skill fade-out arises and how it can be mitigated. An additional benefit of such knowledge is that we deepen our understanding of factors that impact long-term achievement, which has proven difficult due to its high causal density.

[1] Reviews and summaries include Campbell et al. (2001), Barnett (2011), Duncan and Magnuson (2013), Gibbs et al. (2013), Elango et al. (2015), Abenavoli (2019), Bailey et al. (2020), Bruhn and Emick (2023).

Our contribution in this study is to provide a causal understanding of the fade-out of early childhood education programs. We implement a field experiment with two stages of randomization that allow us to identify both the impact of preschool on future achievement, and the effect of having classmates who went to preschool. We begin with a sample of children of low-income families who are invited to attend an early childhood program in the south suburb of Chicago, Chicago Early Childhood Center (CHECC). In CHECC, some children are randomly assigned to receive instructional preschool and some to a control group. This is vital for identifying the causal effect of preschool, given that preschool attendance is generally not distributed as-good-as-randomly in naturally occurring settings. Second, we partner with the local school district to randomly assign these children to classrooms throughout their elementary school years.[2] Unlike settings where peer groups are endogenously formed, such as friendship networks or school classrooms that track students based on ability, this exogenous group formation ensures that classmate traits are conditionally independent of student traits.

Our results fall into three broad areas. First, consistent with past work, we find that CHECC preschool improves cognitive skills in the short-term (Fryer et al., 2020; Castillo et al., 2020). Second, when combining these first stage estimates with rich administrative data drawn from our second stage randomization, we find that preschool students greatly benefit from preschool classmates. In fact, when preschool students are assigned to classrooms with above-median exposure to preschool classmates, they continue to have cognitive skills that are significantly higher than control students. Alternatively, when preschool students are assigned to classrooms with lower exposure to preschool classmates, they are statistically indistinguishable from control students. Thus, while the fade-out effect is present in our data, it can be attenuated with an optimal mix of peers.

Our third area of findings leverages rich survey and administrative data from Chicago Heights school district to explore mechanisms. We find that on average across all district students, exposure to CHECC preschool classmates does not significantly affect students' cognitive skills. This insight suggests that the attenuation of the fade-out effect is social in nature. Using our records of preschool classroom assignments during the preschool experiment, we find that the attenuation of fade-out is driven by elementary school classmates who were assigned to the same preschool classroom. This reveals the importance of social networks even from a young age.

In addition to providing evidence of the social mechanism, we are able rule out other

[2]This does not hold for children who were identified to have special education or health needs, and so we remove these students from our analyses.

potential mechanisms that are often raised in the literature. Two key mechanisms typically discussed are the level and mis-match of peers' skills. Preschool students may fail to retain the gains in cognitive skills when surrounded by non-preschool classmates because these classmates are (1) lower ability, and (2) teachers are unable to adjust instruction to match the preschool student's (higher) ability. We find little evidence supporting these mechanisms. Adding controls for classmate skills—both in levels and similarity relative to the student—generates little change in our estimates. Following a similar approach, we also rule out that preschool classmates generate a better learning environment through reductions in class disruptions. While preschool reduces disciplinary infractions in later years (Castillo et al., 2020), this does not explain the differences in fade-out.

We view our results as potentially drawing interest from several disparate groups. For policymakers, this paper holds important implications for optimal policy in early childhood and program implementation more broadly. We show that long-term program efficacy can hinge crucially on the environments that program participants face in later years. Our exploration of mechanisms reveals ways in which program implementation may take most advantage of the drivers of the synergies we report. In this manner, our results highlight a paradoxical insight: while ardent critics have used the fade-out effect as exhibit A against early childhood programs, our results show that a policy of early education for all children not only can work, but can also be optimal from a human capital accumulation perspective.

Our research also speaks to several bodies of the academic literature. First, our two stages of randomization address key confounds present in work investigating the drivers of fade-out in ECE. This work has been largely correlational, examining key issues such as how skills relate to post-preschool educational inputs such as elementary school curricula (Jenkins et al., 2018) and class size (Magnuson et al., 2007). Directly connected to the mechanisms explored in our study, Botvin et al. (2024) examine the importance of preschool classmates in fade-out by using naturally occurring variation in elementary school classroom assignments following a randomized preschool math intervention. The authors find suggestive evidence that among students assigned to the intervention, those in classes with a greater proportion of classmates who were also assigned to the intervention had higher skills. Whether this relationship is causal, however, remains unclear because endogenous sorting patterns into preschools and elementary school classrooms can also drive them.[3] We tackle such confounds using two stages of randomization, which ensures that our estimates capture the causal link between

[3]For instance, if preschool improves skill gains and elementary schools group together students who are likely to have high skill gains, then it is unclear whether it is being in a class with other preschool classmates or simply the sorting mechanism in into elementary school classes that explains differences in skill growth.

a student's program participation, peers' program participation, and future achievements.

Second, our examination of classmate composition speaks to the large literature on peer effects (Sacerdote (2014) gives a review). Most related to the current paper is Burchinal et al. (2023), who examine the correlation between a student's skills and exposure during elementary school to classmates who previously attended preschool, and Neidell and Waldfogel (2010), who use naturally occurring data on preschool attendance and elementary school classroom composition to document a positive association between student skills and the number of classmates who chose to attend a preschool in earlier years. Our work differs from this literature by utilizing two stages of randomization, which addresses the notion that preschool attendance is not distributed as-good-as-randomly in the naturally occurring settings and can be determined endogenously within peer networks.[4]

Third, this paper connects the literature on optimal policy design and long-term program efficacy with work on peer effects. Social networks have been shown to play an important role in almost every dimension of human behavior, including educational investments (e.g. Joensen and Nielsen, 2018; Bursztyn et al., 2019), educational spillovers within neighborhoods and social relationships (e.g. List et al., 2020; Guo et al., 2024), financial decisions (e.g. Duflo and Saez, 2002; Beshears et al., 2015), health-promoting behaviors (e.g. Duncan et al., 2005; Babcock and Hartman, 2010), on-the-job performance (e.g. Guryan et al., 2009; Mas and Moretti, 2009), and even in highly consequential settings such as policing (e.g. Rivera, 2022; Holz et al., 2023). Dahl et al. (2014) show that take-up of paid paternity leave is largely affected by peer choices of take-up, and that due to a cascade of social influences, long-term take-up rates are substantially higher than when accounting for direct incentives alone. Our results highlight a similar phenomenon, but in the efficacy of program impacts rather than the spread of take-up. In this spirit, while our paper focuses on early childhood education, our results shed light on a more general insight: social interactions can be an important lever for maximizing the persistence of program impacts.

Finally, our work speaks to the recent literature exploring the scale up of public policies (List, 2022; Mobarak, 2022; List, 2024). This work includes general theoretical reasons why policies might not scale (Al-Ubaydli et al., 2020) and empirical work exploring facts around scaling. For example, Agostinelli et al. (2023) study an educational program in Mexico under different modalities to examine whether a high-quality home visitation program can

[4] For instance, if high-income families are more likely to send their children to preschool, then it is unclear whether preschool classmates are beneficial because of preschool status itself or other correlates such as family income. Our design isolates the causal effect of preschool classmates on subsequent skills by randomly assigning preschool to children in early ages, and classmates in later years.

scale effectively. Wang and Yang (2021) analyze over 600 policy experiments in China since the 1980s to examine the efficacy of successful local policies at scale. Finally, Larroucau et al. (2024) examine a multi-year collaboration with policymakers to evaluate information policies implemented at scale to improve students' outcomes in Chile. Collectively, this research paints a rich empirical fabric tied to theory that permits a deeper understanding of the science of using science. Our work takes this literature in a different direction by examining how the number of treated individuals affects program efficacy, highlighting the import of educational network externalities when examining the efficacy of policies at scale.

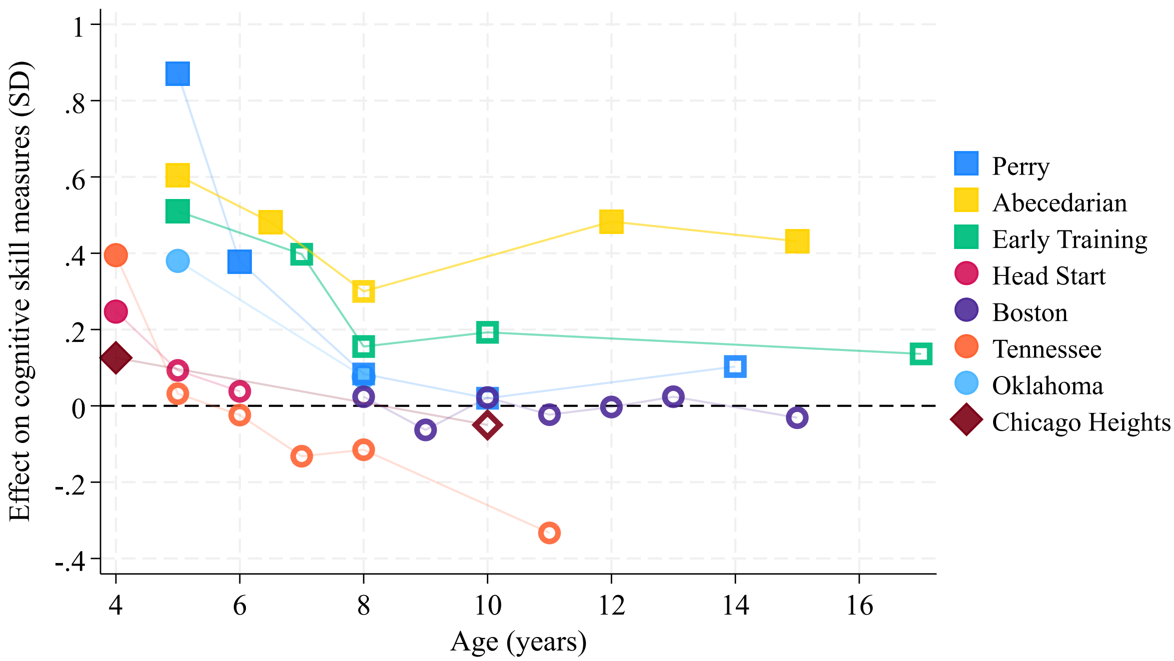
The remainder of our study proceeds as follows. The next section provides further background of ECE fade-out and our field experiment. Section 3 summarizes our empirical framework. Section 4 provides a description of the empirical results. Section 5 concludes.

2 Early Education Background and CHECC

A large body of work has examined the impacts of key early childhood demonstration programs, including the Perry Preschool Project (Weikart, 1970; Schweinhart et al., 1993), Carolina Abecedarian Project (Campbell et al., 2001), and Early Training Project (Gray and Klaus, 1970). These works generally find that the programs—with the exception of Carolina Abecedarian Project—improved child cognitive skills both during implementation and immediate conclusion of the program, but that these effects substantially dissipate by the time the child reaches formal schooling ages (Anderson, 2008; Elango et al., 2015). Studies draw similar conclusions from Head Start (e.g. Deming, 2009; Puma et al., 2012), a federally funded preschool program launched by the United States in 1965 that is the largest early childhood education program in the US to date, as well as more local universal early childhood programs across the country (Bruhn and Emick, 2023) including Boston (Gray-Lobe et al., 2023) and the states of Tennessee (Lipsey et al., 2018; Durkin et al., 2022) and Oklahoma (Gormley Jr et al., 2005; Fitzpatrick, 2008).

We provide ocular evidence of this trend in (Figure 1). There is a burgeoning ECE literature attempting to understand the causes and consequences of the trends in (Figure 1). For example, some studies have examined potential mediators in program effects using variation in subsequent choices, such as how programs affect choice of subsequent school quality (Lee and Loeb, 1995; Currie and Thomas, 2000). Recent work uses state-level court rulings to test whether increased public school funding reinforces the impact of early childhood programs on academic achievements, but finds mixed results (Jenkins et al., 2024; Johnson, 2024). Others theorize that fade-out may be explained by a mismatch in classmate skills, because teachers focus on teaching to the lowest-ability children in the classroom and consequently under-invest in high-achieving preschool students (Duncan and Magnuson, 2013). Our empirical work below complements this work.

Figure 1: Cognitive skill fade-out in early childhood programs



Notes: This figure shows the effects on academic achievements documented from past work evaluating demonstration programs, Head Start (Deming, 2009; Kline and Walters, 2016), and local universal preschool programs, following the spirit of Bruhn and Emick (2023). Colored-in symbols correspond to effects that are significantly different from 0 at the 95% confidence level, while hollow symbols correspond to ones that are not. The demonstration programs are: Perry Preschool Project (Anderson, 2008; Elango et al., 2015), Carolina Abecedarian Project (Anderson, 2008; Elango et al., 2015), Early Training Project (Anderson, 2008; Elango et al., 2015). The local universal programs are Boston (Gray-Lobe et al., 2023), Tennessee Voluntary Pre-kindergarten Program (Lipsey et al., 2018; Durkin et al., 2022), Oklahoma (Gormley Jr et al., 2005; Fitzpatrick, 2008). We additionally include the preschool studied in this paper, Chicago Heights Early Childhood Center (Castillo et al., 2020). All outcomes are cognitive skill measures (IQ or test scores) in standard deviation units. Please see Appendix A for details.

Trends in Figure 1 also raise a related question. Why does the Abecedarian program have long-lasting impacts on cognitive skill while many other programs did not? The reasons are highly speculated in the literature: baseline sample characteristic differences, its longer duration (from birth to age five) and higher intensity through its additional parental and health components (Campbell et al., 2001), participants' access to higher quality schools (Abenavoli, 2019). Our empirical work described below highlights a new potential mechanism: exposure to past program participants. Essentially all of the Abecedarian study sample lived in the same town and attended the same public school system (Campbell et al., 2001) and continued to interact with each other, so much so that the research team could implement a cross-randomized educational support intervention during schooling years and summer camp (Ramey et al., 2000).

2.1 Preschool experiment (CHECC)

Our study begins with the implementation of the Chicago Heights Early Childhood Center (CHECC), a large-scale field experiment conducted during the years of 2010 through 2013 through a partnership with the Chicago Heights school district (Fryer et al., 2020). The center was located in the urban school district of Chicago Heights, Illinois, in the south side of Chicago, where the population in 2010 was nearly 80% Black or Hispanic (U.S. Census Bureau, 2010). The population is relatively low-income, with an average per-capita income of $17,546 and a high school graduation rate of 47%.

Any household in the Chicago Heights school district and neighboring districts with a child of age 3-4 years was eligible to participate in CHECC,[5] and households were recruited through various marketing campaigns (Fryer et al., 2020). Upon signing up, households were randomly allocated to one of four programs: (1) Preschool which provided full-day preschool, (2) Parent Academy which provided incentivized parental education for parents to learn how to teach children at home, (3) Kinderprep which was a shortened summer preschool, and (4) a control group that did not receive any educational services from our research team. In this paper, we focus on the Preschool treatment.

2.2 Elementary school experiment

Concurrent with the implementation of CHECC, we partnered with Chicago Heights school district, which agreed to randomly allocate students across classrooms during elementary school. This meant a partnership with 9 elementary schools for over 6 years. We implemented a completely randomized experiment, so that given a school-grade-year, students were randomly assigned to classrooms.[6] The exception was for students with special education or health needs. This randomization does not hold in later grades, where classes for a given subject (e.g. math) could differentiate based on levels.

[5] Only children who had been identified as needing special education were excluded from CHECC, given staffing limitations.

[6] Other potential designs include varying saturation levels, such as Crépon et al. (2013). We discuss this further in Section 4.3.

During elementary school, students had the same classmates across various subjects. Indeed, over 98% of our student-year observations have the same number of treated classmates in both math and reading classes.

2.3 Data

We combine two main data sources. First, our research team implemented various surveys and assessments while children were participating in CHECC. This includes survey measures of family and parental background, along with assessments of cognitive (such as ability in math, writing, and receptive vocabulary) skills for the child.

We then link the data generated from CHECC with rich administrative data that we receive from Chicago Heights school district. This includes standardized math and reading test scores for grades 2-8, course grades for grades K-8, disciplinary and attendance records for grades K-8. Importantly, we receive report cards that contain classroom identifiers, which allow us to link students with their classmates from grades K-8. We also receive teacher identifiers, which allow us to verify that observed effects are not driven by assignment to particular teachers.

Given that the discussion on fade-out is typically centered around cognitive skills, we focus on standardized test scores and grade point average (GPA). The data contain two types of standardized test scores in math and reading: Measures of Academic Progress (MAP), which are conducted three times per year (once per season: fall, winter, spring) and the Illinois Assessment of Readiness (IAR), which was formerly the Partnership for Assessment of Readiness for College and Careers (PARCC), which is conducted once per year. This means that for each student, we observe at most four test scores per subject per year. We first normalize each test-season to have a mean of zero and standard deviation of one in each grade-year. We then take the average over the four test scores for each subject, so that we have one math test score and one reading test score for each student-year. Please refer to Appendix B for more further details on these data.

2.4 Measuring skills

To focus on fade-out, we investigate the effects of Preschool and classmate composition on cognitive skills. The typical challenge in studying skill development is that latent skills are unobserved, and data instead contain various noisy measures that correspond to an underlying skill. We follow the previous literature and implement factor analysis to create a low-dimensional skill measure from these various measures (e.g. Cunha et al., 2010; Heckman et al., 2013; Attanasio et al., 2020; Joensen et al., 2022). Let $\theta_{i,t}$ be the latent cognitive skill of student i in year t. We do not observe $\theta_{i,t}$, but instead observe M noisy measures as proxies. We use 3 measures of cognitive skill per year: math test score, reading test score, and GPA. We assume a measurement system, so that each measure $m \in \{1, .., M\}$ in the data is a function of latent skill:

$$Z_{i,t}^{m} = \mu_t^m + \lambda_t^m \theta_{i,t} + \eta_{i,t}^m \tag{1}$$

where $Z_{i,t}^m$ is an observed measure and $\eta_{i,t}^m$ is the corresponding error term. $\lambda_{i,t}^m$ is the measurement factor loading, and μ_t^m the factor intercept. Both are parameters to be estimated, and can be identified from the data. We estimate the latent skill factor using Bartlett scores (Heckman et al., 2013). Please see Appendix C for details. We use this cognitive skill factor as the outcome of interest in the main text, and show analyses for each component in the Appendix.

2.5 Sample construction

We examine the impact of classmate composition on academic achievement from Kindergarten through grade 5. We stop at grade 5 because classroom assignment in later grades can depend on achievement levels (e.g. honors math versus general math).

We drop classrooms with fewer than 10 students, because they are typically classes for students with special education or health needs, and are not subject to random assignment. This corresponds to 3 percent of the student-year observations. We additionally drop students who move schools or classes during the year, since we do not observe move dates and therefore do not observe how long a student was exposed to each set of classmates. This corresponds to less than 2 percent of student-year observations in our sample.

When combining data from CHECC with administrative data from Chicago Heights school district, we subset to the school-grade-years that contain at least one CHECC student. This choice makes little impact on our results.

2.6 Summary statistics

Summary statistics for the elementary school sample are contained in Table 1. Of the 2,208 children who participated in CHECC, and were assigned a treatment status, we observe 867 in

Chicago Heights. Of those, 203 were assigned to Preschool and 440 to control.[7] As discussed by Castillo et al. (2020), it is reasonable to observe students outside of Chicago Heights school district, particularly given that CHECC purposefully admitted households located outside of the district. Our sample additionally contains 2,435 students who attended a school in the district but were not associated with CHECC, meaning that they were neither in control nor any treatment group.

Table 1: Summary Statistics

	Mean	SD
Demographics		
Race: White	0.39	(0.49)
Race: Black	0.55	(0.50)
Hispanic: Yes	0.51	(0.50)
Gender: Female	0.49	(0.50)
CHECC: participated	0.26	(0.44)
CHECC: participated and assigned Preschool	0.06	(0.24)
Classmate Traits		
N total	18.78	(4.85)
N assigned control in CHECC	3.45	(2.29)
N assigned Preschool in CHECC	1.67	(1.61)
At least one assigned Preschool in CHECC	0.66	(0.47)
Elementary School Outcomes		
GPA (out of 4)	2.60	(0.97)
MAP math (national percentile rank)	36.86	(24.08)
MAP reading (national percentile rank)	42.91	(26.95)
MAP math (rit)	196.83	(16.16)
MAP reading (rit)	193.71	(17.27)
PARCC math (score)	719.92	(27.68)
PARCC reading (score)	723.34	(32.98)
Has disciplinary infractions	0.19	(0.40)
Number of disciplinary infractions	0.61	(2.65)
Days Absent	10.36	(9.64)
Days Absent (fraction of total)	0.06	(0.06)
N Classrooms	574	
N Students	3304	
N Student-Year	11473	

Notes: This table shows the demographic characteristics, classmate traits, and elementary school outcomes for the analysis sample. Demographics are means at the student-level. Classmate traits and elementary school outcomes are at the student-year level.

As can be seen in Table 1, students in a given year had on average of roughly 19 class-

[7] The rest of the students were assigned to one of the remaining treatment groups, as explained in Section 2.1.

mates, with around 1.5 of these classmates being randomly assigned to CHECC Preschool. In addition, 66 percent of student-year observations were associated with at least one Preschool classmate. Importantly, standard deviations in Table 1 provide optimism that there is enough variation in classmate composition to explore whether that can be a key factor potentially attenuating the fade-out effect.

3 Empirical Framework

We first describe our parameter of interest and our identification assumptions. Let D be an indicator for assignment to Preschool. We omit student-level subscripts for simplicity. Let N denote the exposure to Preschool classmates. For simplicity, suppose that exposure to Preschool classmates is binary. Let $Y_{D,N}$ be the potential outcome; for example, cognitive skills in 3rd grade, under exposure N to Preschool classmates and Preschool assignment D. Let Y denote the actual observed outcome. When N is binary, Y can be expressed as a function of the four potential outcomes:

$$Y = Y_{1,1}DN + Y_{1,0}D(1-N) + Y_{0,1}(1-D)N + Y_{0,0}(1-D)(1-N) \tag{2}$$

Our goal is to identify the extent to which exposure to Preschool classmates moderates the impacts of Preschool on academic achievement:

$$\delta \equiv \mathbb{E}\left[\left(\underbrace{Y_{D=1,N=1} - Y_{D=1,N=0}}_{\text{effect of Preschool when high exposure}}\right) - \left(\underbrace{Y_{D=0,N=1} - Y_{D=0,N=0}}_{\text{effect of Preschool when low exposure}}\right)\right] \tag{3}$$

The main challenge is that only one of the four potential outcomes is observed per student. We rely on three identification assumptions to overcome this challenge. We first state our assumptions, and then provide evidence for them in the next section. We focus our attention on the population that CHECC drew its sample from.

Assumption 1. *Preschool assignment is independent of potential outcomes*

$$D \perp Y_{d,n} \;\forall\; d, n \tag{4}$$

This assumption yields $\mathbb{E}[Y_{d,N=1} - Y_{d,N=0}] = \mathbb{E}[Y_{D,N=1} - Y_{D,N=0} | D = d] \;\forall d$. Our second assumption arises from the fact that we only observe exposure to Preschool classmates for children who enrolled in Chicago Heights.

Assumption 2. *The choice to attend Chicago Heights school district, and the choice of school-grade-year is independent of the potential outcomes and unaffected by Preschool assignment*

$$\gamma \perp Y_{d,n}, D \ \forall \ d, n \tag{5}$$

where γ represents a student's school-grade at a particular year. This assumption provides $\mathbb{E}[Y_{D,N=1} - Y_{D,N=0}|D = d] = \mathbb{E}[Y_{D,N=1} - Y_{D,N=0}|D = d, \gamma] \ \forall d, \gamma$. Finally, we require an assumption on the allocation of exposure to Preschool classmates.

Assumption 3. *Exposure to Preschool classmates is independent of potential outcomes, conditional on school-grade-year*

$$N \perp Y_{d,n} \mid \gamma \ \ \forall \ d, n \tag{6}$$

With these three assumptions, we are able to rewrite the parameter of interest in terms of conditional means that are observed in the data (see Appendix D for details):

$$\begin{aligned} \delta = \mathbb{E}[Y|D = 1, N = 1, \gamma] - \mathbb{E}[Y|D = 1, N = 0, \gamma] \\ - \mathbb{E}[Y|D = 0, N = 1, \gamma] + \mathbb{E}[Y|D = 0, N = 0, \gamma] \end{aligned} \tag{7}$$

3.1 Assessing the identifying assumptions

To satisfy the first identifying assumption, we rely on random assignment in our Preschool experiment at the Chicago Heights Early Childhood Center (CHECC). Castillo et al. (2020) and Cappelen et al. (2020) show that pre-CHECC covariates are balanced across treatment groups. Table A2 confirms that pre-CHECC covariates are also balanced across the treatment groups among children we later observe in Chicago Heights school district.

Our second identifying assumption is violated if random assignment to Preschool causes differential migration out of the Chicago Heights district. We compare enrollment patterns across treatments and find little evidence that Preschool affected students' choices of attending Chicago Heights school district, or a particular school in Chicago Heights (Table A3).

Finally, our third identifying assumption is satisfied by conditional random assignment of elementary school students to classrooms. We conduct placebo tests, testing for significant relationships between pre-classroom-assignment characteristics and classroom assignment, conditional on school-grade-year. We estimate for student i in grade g during year t at school s:

$$Y_{i,t-1} = \pi^0 + \pi_1 \mathbf{N^R_{i,t}} + \pi_2 \tilde{N}^R_{i,t} + \lambda^{Ntotal}_{i,t} + \gamma_{t,g(i,t),s(i,t)} + \nu_{i,t} \tag{8}$$

for $R \in \{\text{Preschool, control, non-CHECC}\}$, where $Y_{i,t-1}$ is a trait of i that is determined before time t (such as GPA the year prior). $\tilde{N}^R_{i,t}$ is the total number of R-type students who enroll in grade g at school s in year t, besides student i. As described in Guryan et al. (2009), including this control is important as it accounts for the mechanical differences in group means that arise because an individual cannot be his or her own classmate. Table A4 shows our estimates of π_1. We find no evidence of significant correlation between pre-determined characteristics and classroom assignment.

3.2 Empirical approach

In this section, we highlight two common obstacles in the peer effects literature and how we overcome each, in turn. First, a primary difficulty in identifying peer effects is the reflection problem (Manski, 1993): peers may affect a student's outcome, but the student may simultaneously impact peer outcomes. Correlations in outcomes therefore confound the effect of peers with the effect onto peers. Our approach addresses this potential problem by examining a classmate trait that is determined before elementary school and randomly assigned—whether the classmate was randomly assigned to Preschool during CHECC—so that a student's academic achievement in elementary school cannot influence the classmate trait of interest.

Second, peer groups are often endogenously formed, so that any correlation between peer outcomes can confound peer effects with the notion that individuals of similar outcomes are more likely to be in the same peer group. For instance, Preschool students may often be in the same elementary school classrooms if Preschool improves skills and schools assign classrooms based on ability. In this case, it appears that being surrounded by Preschool classmates increases skills, but the correlation is driven by a causal path moving in the opposite direction. Past work in the education context has typically tackled this by using either naturally occurring, across-cohort, within-school variation in student composition (e.g. Bifulco et al., 2011; Lavy and Schlosser, 2011), or random assignment to peer groups (e.g. Carrell et al., 2009; Booij et al., 2017). Our design approach follows the latter literature, where students are randomly assigned to classrooms.

Specifically, we test how changes in classmate composition in Preschool status affect outcomes. Our specification mimics List et al. (2020), who estimate spillovers of CHECC on neighbors during its implementation years, and other work, such as Miguel and Kremer (2004). We estimate, for student i in grade g during year t in school s:

$$
\begin{aligned}
Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 \mathbf{N}_{\mathbf{i,t}}^{\mathbf{Preschool}} + \beta_3 \mathbf{N}_{\mathbf{i,t}}^{\mathbf{Preschool}} \times D_i \\
+ \beta_4 N'_{i,t} + \beta_5 N'_{i,t} \times D_i + \beta_6 X_{i,t} + \lambda_{i,t}^{Ntotal} + \gamma_{t,g(i,t),s(i,t)} + \varepsilon_{i,t}
\end{aligned}
\tag{9}
$$

where $Y_{i,t}$ is i's cognitive skill in year t, D_i is a vector of pre-elementary school indicators for whether the child was assigned to Preschool (and takes on value 0 if the child did not participate in CHECC), one of the other CHECC programs besides control, or did not participate in CHECC. $\mathbf{N}_{\mathbf{i,t}}^{\mathbf{Preschool}}$ is the number of classmates who were assigned to Preschool, and $N'_{i,t}$ is a vector containing the numbers of classmates who: did not participate in CHECC, were assigned to the CHECC parental program, or were assigned to the CHECC summer program.

The omitted category in this estimation is the number of classmates who participated in CHECC and were assigned to the control group. $X_{i,t}$ is a vector of controls for race (Black or other or missing), ethnicity (Hispanic or missing), gender. $\lambda_{i,t}^{Ntotal}$ denotes fixed effects for the total number of classmates, and $\gamma_{t,g(i,t),s(i,t)}$ represents school-by-grade-by-year fixed effects. β_3 captures the difference between Preschool and control students in skill gains caused by assignment to more Preschool classmates, relative to more control classmates. In Section 4.3, we consider alternative specifications beyond a linear count of peers.

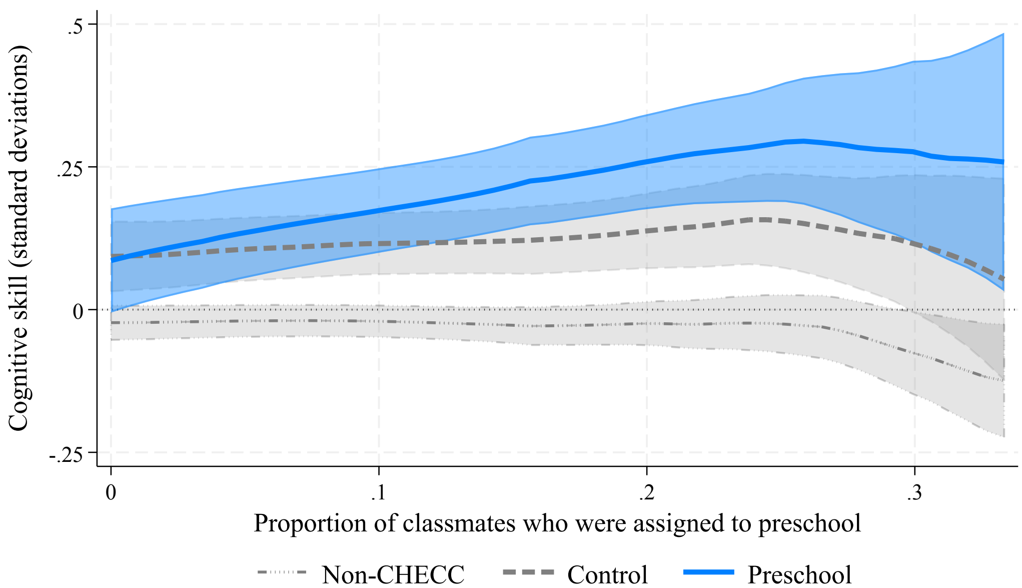
4 Field Experimental Results

This section provides a summary of our experimental results. We begin by exploring the effects of classmate exposure levels and examine tests of heterogeneity. We then discuss the implications of our fade-out results and various robustness tests. We conclude this section with a discussion of potential mechanisms.

4.1 Effects of classmates who were assigned to preschool

To consider the potential effects of having classmates who were randomly assigned to Preschool, we first show the unconditional relationship between a student's exposure to Preschool classmates and cognitive skills at the end of the school-year (Figure 2). The figure contains students during grades 2-5, since test scores are observed starting from 2nd grade. The figure shows results for three student types, those who: i) did not participate in CHECC (gray dotted line), ii) participated and were randomly assigned to control (gray dashed line), iii) participated and were randomly assigned to Preschool (blue solid line). We fit the three lines using a kernel-weighted local polynomial regression, and we shade areas corresponding to 90 percent confidence intervals. The slope of each line shows the unconditional association between a student's exposure to Preschool classmates and the student's cognitive skills. Interestingly, students randomly assigned to Preschool appear to have a steeper slope than the other two groups, suggesting that Preschool children benefit more from Preschool classmates than students who were control or did not attend CHECC.

Figure 2: Unconditional relationship between cognitive skill and classmate composition



Notes: This figure shows the relationship between a student's cognitive skills and the proportion of the student's classmates who were assigned to Preschool, by whether the student did not participate in CHECC (gray dotted line), participated and was randomly assigned to control (gray dashed line), or participated and was randomly assigned to Preschool (blue solid line). The sample contains student-year observations during grades 2 through 5. Lines are fit using a kernel-weighted local polynomial regression. Shaded areas correspond to 90% confidence intervals.

We confirm these results using our main specification that includes various controls– importantly, school-by-grade-by-year fixed effects (Equation 9). For the effects of Kindergarten and 1st grade classmates, when test scores are not observed for that year, we estimate Equation 9, except replacing the school-by-grade-by-year fixed effects with school-by-year given that we explore classroom composition for two two grades separately. We verify that classmate composition is not associated with differential attrition out of the sample in subsequent grades (Table A5). Table 2 summarizes our main empirical results.

Table 2: Effects of elementary school classmates who were assigned to Preschool

	Classmates during...					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.05*	-0.05	0.01	-0.02	0.02	-0.01
	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
N Preschool classmates × Preschool		0.04		0.08*		0.10***
		(0.05)		(0.04)		(0.04)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4118	4118	4475	4475	5688	5688
N Students	1446	1446	1575	1575	2237	2237
Adjusted R^2	0.23	0.23	0.23	0.23	0.18	0.19

Notes: This table shows the effect of having elementary school classmates during (columns 1-2) Kindergarten or (columns 3-4) 1st grade or (columns 5-6) grades 2-5 who were randomly assigned to Preschool, on cognitive skills during grades 2-5. Dependent variable for all columns is cognitive skill. Sample contains student-year observations during grades 2-5. Each column follows Equation 9, except for columns 1-4 which do not include fixed effects for grades given that they focus on classmate composition from a single grade. "N Preschool classmates" is the number of classmates during the corresponding grade who were randomly assigned to Preschool. School-by-grade-by-year fixed effects correspond to the school and year of the grade indicated in the column header. Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to Preschool. For additional precision, we include measures collected from CHECC (age at experiment, birth weight, mother's education, mother's age at child's birth, family income, indicator for two-parent household, pre-elementary school cognitive and executive function skills), and indicators for each on whether the student is missing an observation for the indicator. Students who were not part of CHECC are imputed as the sample mean value. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Table 2 reveals that when averaging across all students in the district, having more Preschool classmates relative to non-Preschool ones (i.e. those in control or those who did not participate in CHECC) does not significantly impact cognitive skills. However, this average masks an important heterogeneity. Preschool assignment is synergistic across participants later on in life: Preschool children benefit greatly from having other Preschool children in the same class. Holding constant the total number of students in class, one additional Preschool classmate significantly increases cognitive skills by roughly 0.10 SD. Our estimates remain largely unchanged even with the inclusion of teacher fixed effects, confirming that our results are not driven by assignment to particular teachers.

These gains are driven by improvements in all three cognitive skill components: math test score, reading test score, and GPA (Table A6). For MAP math and reading standardized tests, we also observe the national percentile rank and RIT score. Using these metrics, we find that national percentile ranks (Table A7) and RIT scores (Table A8) lead to similar conclusions as the main analysis. Taken together, our evidence suggests that the synergistic role of Preschool classmates is not simply an artifact of how cognitive skills are measured or scaled (Cascio and Staiger, 2012; Wan et al., 2021).

While columns 4 and 6 of Table 2 show that classmates during grades 1 through 5 play an important role in mitigating fade-out, column 2 shows that we fail to detect this effect in Kindergarten. Interestingly, Botvin et al. (2024) also find this qualitative difference in the effects of Kindergarten and later-grade classmates using naturally occurring variation in elementary school classroom assignments following a randomized preschool math intervention. There are several potential reasons why we find this pattern in our data. First, our outcome variable (cognitive skills) is only observed starting from grade 2, so any effects that occur in Kindergarten may be attenuated due to changes in classmate composition in subsequent grades. Second, the social and collaborative nature between classmates, as well as its substitutability with family involvement, may evolve over grades. We view this area as ripe for future exploration.

4.1.1 Heterogeneity

Our data are rich enough to explore various types of heterogeneity. We find that our Preschool classmate synergy result is driven by students whose mothers have an education level no greater than high school rather than students whose mothers completed higher levels of education (Table A9). This suggests that Preschool classmates can, in part, operate as substitutes for parental education, though the difference-in-difference is not statistically significant. Point estimates also suggest that the observed synergy is higher for boys and Hispanic students relative to girls and non-Hispanic students, which is similar to List et al. (2020) who find across-neighbor spillovers in cognitive skill improvements caused by Preschool are driven by Hispanic families. However, we fail to reject that the effect is different across student gender, ethnicity, and race. Future work should explore these factors using causal moderation to pin down exact sources of heterogeneity to provide advice to policymakers on viable personalized programs (List, 2025).

4.2 Implications for fade-out

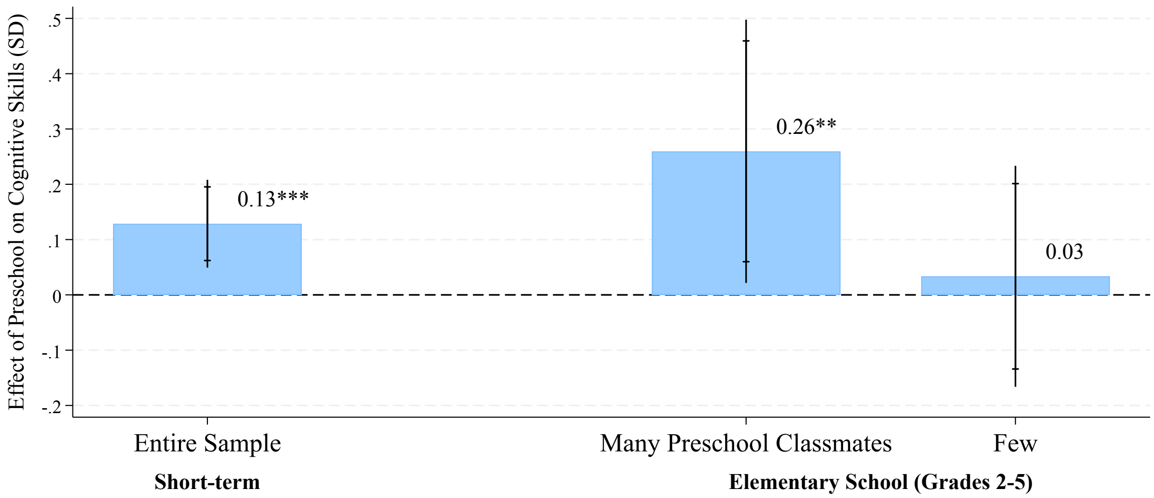
The previous results show that Preschool students have higher cognitive skills when surrounded by other Preschool students. In this section, we interpret these results in the context of Preschool treatment effects. We summarize our results in Figure 3. To construct the figure, we first show the effect of Preschool on cognitive skills immediately after program implementation (leftmost bar). We then show the effect several years later when the children are in elementary school (two rightmost bars). The elementary school bars separate the effect into student-year observations with high exposure to Preschool classmates and those with low exposure. High exposure is a binary variable that is constructed by residualizing school-grade-year fixed effects and the number of classmates who were neither Preschool nor control (Parent Academy, Kinderprep, or did not participate in CHECC, a variable for each) from the number of Preschool classmates, and then binning this residual by its median value. We then regress a student's test scores in a given year on Preschool assignment interacted with exposure to estimate the impacts of Preschool separately by exposure group.[8]

Several results can be gleaned from Figure 3. First, as shown by Castillo et al. (2020), immediately after program implementation, Preschool CHECC children had 0.13 standard deviations higher cognitive skills than control children (leftmost bar of Figure 3). Second, turning to the impacts of Preschool on cognitive skills during elementary school, the two rightmost bars of Figure 3 reveal that during grades 2-5, when Preschool children are exposed to many Preschool classmates, they have roughly 0.26 standard deviations higher cognitive skills than control children. Returning to Figure 1, this magnitude is most similar to the elementary school impacts of the Early Training Program (Anderson, 2008; Elango et al., 2015). In contrast, when Preschool children are surrounded with very few Preschool classmates, they are statistically indistinguishable from control children. This result shows that classroom peers can not only attenuate the fade-out effect, but reverse it, leading to even larger early education effects in the long-run.

We next go beyond means and explore how classmates affect the skill distribution. Figure 4 shows the cognitive skill distribution at the student-year level during grades 2-5, separated out by a CHECC student's CHECC treatment status and exposure to Preschool classmates that year. We use the same exposure categories as in Figure 3, and residualize out school-grade-year fixed effects and CHECC experiment controls from cognitive skills. Figure 4a confirms the prior finding that among Preschool students, students had higher

[8]To account for attrition out of the sample, we follow Castillo et al. (2020) and use inverse probability weighting. In practice, the results are essentially unchanged whether we do this or not, consistent with the notion discussed in Section 3.1 that students did not systematically migrate out of the sample.

Figure 3: Effects of Preschool and Preschool classmates on cognitive skills



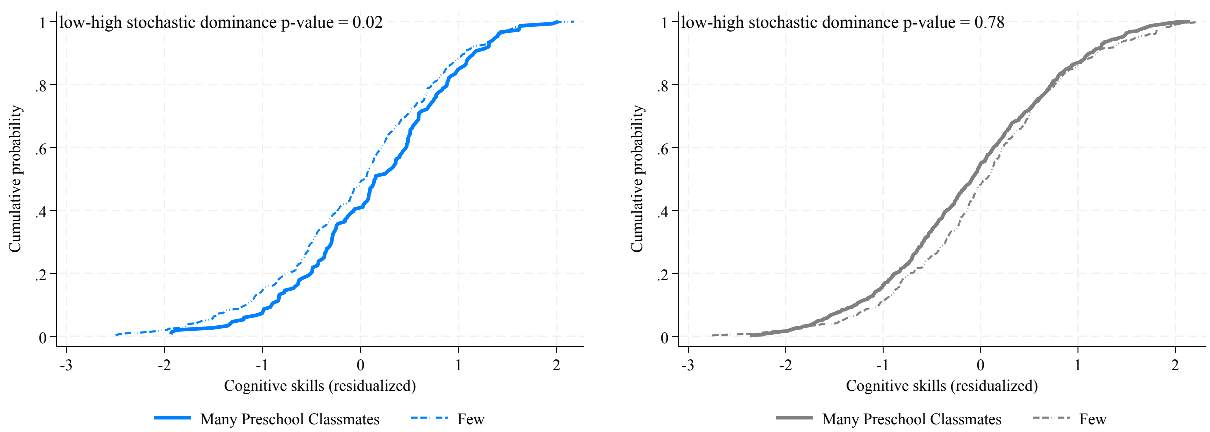
Notes: This figure shows the effect of Preschool on cognitive skills immediately after program implementation (leftmost bar) and several years later when the children are in elementary school (two rightmost bars). The elementary school bars separate the effect into student-year observations who have high exposure to Preschool classmates and those with low exposure. High exposure is a binary variable that is constructed by residualizing out school-grade-year fixed effects and the number of classmates who were not Preschool nor control (Parent Academy, Kinderprep, or did not participate in CHECC, a variable for each) from the number of Preschool classmates, and then binning this residual by its median value. "Many" refers to above-median in this residualized value, and "few" to below-median.

cognitive skills when assigned to a classroom with a relatively high exposure to Preschool classmates. It further reveals that this is not driven by outliers but instead a rightward shift in the entire skill distribution. We implement Barrett and Donald (2003)'s consistent test of first-order stochastic dominance, with the null hypothesis that the skill distribution under high exposure does not stochastically dominate the distribution of skills under low exposure.[9] When subsetting to Preschool students, we reject the null and conclude that the student skill distribution under high Preschool classmate exposure stochastically dominates the distribution under low exposure. Figure 4b shows that this is not the case for CHECC control students. If anything, higher exposure for control students shifts the skill distribution to the left.

[9]We implement Barrett and Donald (2003)'s bootstrap approach with 1000 draws using Schaub and Schaub (2024).

Figure 4: Distributional effects of Preschool classmates on cognitive skills

(a) Preschool CHECC students (b) Control CHECC students



Notes: This figure shows the student-year distribution of cognitive skills by CHECC treatment status (Preschool in 4a and control in 4b) and exposure to Preschool classmates during that year in elementary school. We residualize out school-grade-year fixed effects and CHECC experimental controls from cognitive skills. High exposure is a binary variable that is constructed by residualizing out school-grade-year fixed effects and the number of classmates who were not Preschool nor control (Parent Academy, Kinderprep, or did not participate in CHECC, a variable for each) from the number of Preschool classmates, and then binning this residual by its median value. "Many" refers to above-median in this residualized value, and "few" to below-median. The p-values in the top corner correspond to Barrett and Donald (2003)'s consistent test of first-order stochastic dominance, where the null hypothesis is that the skill distribution for high exposure does not stochastically dominate the low exposure distribution. We implement the test's bootstrap approach with 1000 draws using Schaub and Schaub (2024).

4.3 Robustness

One critique of peer effects studies that use complete random assignment to peer groups is that they rely on finite-sample variation in peer groups, so that traditional inference techniques may not be valid. We follow Athey and Imbens (2017) and construct p-values using randomization-based inference, as implemented in past work on peer effects (e.g. Carrell et al., 2013, 2019; Rivera, 2022). In our context, this method takes the peer network (classroom assignments) as given, and tests the sharp null hypothesis that classmates' Preschool statuses do not have differential effects on Preschool students. Figure A3 shows the distribution of estimated coefficients. As in the main analysis, the differential effect of Preschool classmates on Preschool students is statistically significant at conventional levels ($p = 0.01$).

Finally, we conduct robustness checks to confirm that results are not driven by our particular choice of functional form. The main specification uses a linear term for the number of Preschool classmates, conditional on the total number of classmates (Equation 9). While using fixed effects for the total number of classmates allows us to accommodate some non-linearity, it is still possible that our use of a linear Preschool classmate term imposes restrictions from fitting the true relationship. We first note that a plot of the raw data suggests that the patterns are not driven by outliers (Figure 2), and the finding that Preschool classmates improve Preschool students' skills persists when we bin exposure by above- and below-median (Figure 3). We estimate two additional specifications, one where we use the proportion of classmates who were in Preschool as a regressor and another where we model whether a student has at least one Preschool classmate. In both robustness tests we find the same qualitative conclusion as the main analysis: Preschool students benefit substantially from Preschool classmates (Table A10).

4.4 Mechanisms

We next explore potential mechanisms driving the observed synergies we find between Preschool children and Preschool classmates. Given random assignment during CHECC, we rule out the hypothesis that the effect is due to differences in baseline traits between Preschool and control classmates. For instance, it is not the case that the Preschool interaction effect is driven by differential impacts of classmates from higher-income families, since on average Preschool and control students are balanced on these predetermined characteristics. We confirm this fact by including controls for classmate demographics (race, gender, ethnicity), and find that the empirical estimate of the synergy effect remains unchanged (column 1 of Table 3).

4.4.1 Network formation during preschool

A first mechanism we explore is that Preschool children may benefit from Preschool classmates due to shared past networks, building on past work that has looked at the relationship between enrollment into a new school (e.g. college) and prevalence of peers from the same prior school (Fletcher and Tienda, 2009; Lavy and Sand, 2012; Herbst et al., 2023). For example, students may stay on task or work together in groups within the classroom when surrounded by peers of previously-formed networks. We additionally view this channel as linked to past work on the potential gains from repeated interactions with teachers, or "looping" policies (Hill and Jones, 2018), as they may operate under similar psychological channels such as shared understanding and ease of communication.

To test whether Preschool synergy is attributable to network formation, we estimate

Table 3: Mechanisms driving Preschool synergy in cognitive skill development

	(1)	(2)	(3)	(4)
N Preschool classmates × Preschool	0.10***			0.09***
	(0.04)			(0.04)
N Preschool classmates who were in same Preschool year		0.11***		
		(0.04)		
N Preschool classmates who were in different Preschool year		0.05		
		(0.07)		
N Preschool classmates who were in same Preschool class			0.21*	
			(0.11)	
N Preschool classmates who were not in same class			0.09**	
			(0.04)	
N Neighbors				0.03**
				(0.01)
School × Grade × Year FE	×	×	×	×
Demographics	×	×	×	×
Classmate Demographics	×			
N	5688	5688	5688	5688
N Students	2237	2237	2237	2237
Adjusted R^2	0.20	0.20	0.20	0.20

Notes: This table shows the effect of classmate characteristics on cognitive skills. Sample contains student-year observations during grades 2-5. The variable "N Preschool classmates who were in same Preschool year" is the number of classmates who were assigned to Preschool in the same year as the student. The variable "N Preschool classmates who were in same Preschool class" is the number of classmates who were assigned to Preschool classroom as the student. This same-class count variable in practice is essentially identical to including an indicator for having any classmates who were in the same Preschool class because majority of those who have any have only one. These same-year and same-class variables take on a value 0 if the student was did not participate in CHECC or was not assigned to CHECC Preschool. Classmate demographics include the fraction of classmates associated with each possible gender (female), race (Black, non-White other, or missing), and ethnicity (Hispanic). Demographics include gender, race, ethnicity. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

the additional gains that students receive from having classmates who they were exposed to during Preschool. We explore this potential mechanism using two types of variation. First, using rosters from CHECC, we identify which students attended Preschool during the same years. We find that the synergy between Preschool students is largely driven by elementary school classmates who had attended CHECC Preschool in the same year as each other (column 1 of Table 3).

Second, in the same spirit, we also examine the effects of classmates who were in Preschool and who were assigned to the same CHECC Preschool class. Among student-year observations for those who were assigned to CHECC Preschool, approximately 18 percent had exactly one elementary school classmate who was assigned to the same CHECC Preschool class, and less than one percent of observations had more than one. Point estimates from our regression suggest that the synergy between Preschool classmates during elementary school is greater when the children had been classmates during Preschool (column 3 of Table 3), though we fail to reject the homogeneity null of these two coefficients. This is qualitatively consistent with past work showing that exposure to friends during school can increase test scores, potentially mediated through behaviors such as time spent on homework (Lavy and Sand, 2012).

Additionally, consistent with this notion that having more classmates from similar networks can improve skills, we find small but positive effects of being assigned to a classroom with more neighbors (column 3 of Table 3), defined as students who lived within 1000 kilometers when CHECC was implemented as in (List et al., 2020). This result reinforces the social network effect observed in our data.

4.4.2 Classmate skills

We next turn to the role of peer skills as a potential mechanism. Given that the Preschool caused short-term improvements in skills, we test whether Preschool classmates create a more conducive learning environment for other Preschool children via providing a higher-skilled peer group. To examine this mediation path, we include controls for students' classmate skills. We use classmate skills the year prior to avoid the reflection problem (Manski, 1993), meaning that we subset to classmate composition during grades 3-5 given that test score measures begin at grade 2.

Table 4 summarizes our empirical estimates. The first column estimates the main effect of Preschool classmates on Preschool children, as in Table 2 but now subsetting to grades 3-5. The estimates are consonant with the earlier empirical results on program efficacy. We next examine whether Preschool synergy arises due to improved skill matching. The "boutique," or tracking model of peer effects, asserts that students benefit most when situated with similar-skilled peers. For example, if teachers target lessons to the average, lowest-skilled, or high-level students, then having a greater number of Preschool classmates may be beneficial to Preschool students through increasing similar-skilled classmates. To test this hypothesis, we include the number of classmates who have congruent skills (within 0.1 standard deviations of the student in the prior year).[10] We find that this has negligible impact on future skills, and the coefficient on the Preschool interaction term remains unchanged (column 2 of Table 2). In other words, we find little evidence for the boutique model, and therefore we argue that the evidence is scant that better-targeted teaching is the driving mechanism of our results.

[10]Empirical results are unchanged if we use a different cutoff, such as being within 0.5 standard deviations, to define similar skills.

Table 4: Role of classmate skills in cognitive skill development (grades 3-5)

	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates × Preschool	0.11**	0.11**	0.11**	0.11**	0.11**	0.11**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
N classmates who had similar skill (prior year)		0.01				
		(0.02)				
Average classmate skill (prior year)			0.18**			
			(0.07)			
Min classmate skill (prior year)				0.12**		
				(0.05)		
Max classmate skill (prior year)					0.24***	
					(0.06)	
Weighted classmate skill (prior year)						0.43***
						(0.07)
Skill elasticity of substitution (α)						2.43***
						(0.88)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	3489	3489	3489	3489	3489	3489
N Students	1718	1718	1718	1718	1718	1718
Adjusted R^2	0.11	0.11	0.11	0.11	0.12	0.14

Notes: This table shows the effect of classmate characteristics (past Preschool assignment and skills from the year prior) on cognitive skills. Sample contains student-year observations during grades 3-5, which are the grades for when lagged test scores are available. α is the elasticity of substitution between classmate skills, as explained in Section 4.4.2. Demographics include gender, race, ethnicity. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

We next turn to whether the level of classmate skill is an important mechanism. We start by implementing the functional form that is commonly used in the literature: the "linear-in-means" unweighted average over classmates' cognitive skills the year prior. More specifically, for student i, we include the regressor $\frac{1}{|C_{i,t}|}\sum_{j\in C_{i,t}} Y_{j,t-1}$ in our regression specification (Equation 9), where $C_{i,t}$ is the set of i's classmates in year t, and $Y_{j,t-1}$ is classmate j's skill level in the year prior. As a robustness test, we additionally explore the inclusion of other aspects of the distribution, including the minimum and maximum classmate skill level.

Columns (3-5) of Table 2 summarize these empirical estimates. We find that the coefficient on classmate skills is positive in all specifications, suggesting that on average, being surrounded by higher-skilled classmates improves cognitive skills. However, the coefficient on the interaction between Preschool and number of classmates who were in Preschool remains largely unchanged by the addition of these controls, suggesting that changes in classmate skill level is not the driving mechanism. Yet, one concern with these estimates is that the linear-in-means model is mis-specified and therefore underestimates the role of peer skills.

As a robustness check, we relax the substitutability of classmates in cognitive skill development, allowing for the possibility that classmate skills do not enter the production function of cognitive skills as a simple average. We model peer effects using a constant elasticity of substitution (CES) aggregator, adapting the approach due to Boucher et al. (2024) for outcomes centered around zero. Specifically, we include the regressor $\frac{1}{\alpha}\log(\frac{1}{|C_i|}\sum_{j\in C_i}\exp(Y_{j,t-1})^{\alpha})$ in our model. We then estimate the coefficient on this regressor, whereby α which is the elasticity of substitution across classmates.[11] We estimate the parameters using non-linear least squares.

Interestingly, our estimated elasticity of substitution between classmates' cognitive skills (α) is positive, meaning that classmate cognitive skills operate as substitutes in the production of cognitive skills (column 6 of Table 4).[12] In fact, we reject the linear-in-means model. However, most importantly, we find that even after relaxing substitutability, the coefficient on the interaction of interest remains unchanged by the inclusion of this control. Taken together, the evidence suggests that differences in classmate skills are an unlikely mediator for the impacts of Preschool classmates.

4.4.3 Behavioral disruptions

We close by considering the effects of peers mis-behaving in the classroom. Past work finds that disruptive peers can worsen academic outcomes (Carrell and Hoekstra, 2010; Lavy and Schlosser, 2011). And, importantly, Castillo et al. (2020) find that CHECC Preschool reduced disciplinary infractions. Using data on classmates' disciplinary infractions and absenteeism, we explore whether Preschool classmate synergy can be explained by differences in classmate mis-behaviors, following the same steps as in the above section. Perhaps contrary to intuition, we find little evidence that classmates with more disciplinary infractions or absenteeism worsen cognitive skills (Table A11). We again find virtually no change in our coefficient on the interaction between Preschool and Preschool classmates, suggesting that our synergistic Preschool effects are not driven by changes in classmate behaviors.

[11] Negative values of α correspond to classmate skills operating as complements in cognitive skill development, while positive values correspond to substitutability. This functional form nests the linear-in-means model. The two are equivalent as α tends to zero: $\lim_{\alpha\to 0}\frac{1}{\alpha}\log(\frac{1}{|C_i|}\sum_{j\in C_i}\exp(Y_{j,t-1})^{\alpha}) = \frac{1}{|C_i|}\sum_{j\in C_i}Y_{j,t-1}$.

[12] This is consonant with Boucher et al. (2024) who estimate that peers' GPAs function as substitutes in the production of GPA.

5 Conclusion

Early childhood education sets the stage for lifelong learning and development, making it a critical investment for individuals and society as a whole. Furthermore, quality early childhood education can help bridge the gap between children from different socio-economic backgrounds, promoting equity and social justice. For these purposes, measuring the economic benefits and costs of early childhood education programs has taken on great policy import.

Yet, the evidence thus far does not paint a consistently compelling picture. For example, the fade-out effect–the phenomenon wherein the initial academic gains from early childhood education diminish over time–has been oft-observed. This paper takes this literature in a new direction by using a 2-stage randomization field experiment to examine the fade-out effect and its various potential mechanisms. In particular, we study the role of classmates in the persistence of effects from an early childhood program that we started in 2010. We randomize Preschool offers and elementary school classroom composition, which allows us to identify both the effect of preschool and the effect of Preschool classmates on future skills.

We find that preschool children greatly benefit from being assigned to classrooms with more preschool classmates, holding constant the total number of classmates. Consequently, the impacts of preschool in later years is greatly affected by classmate composition. These results highlight the importance of classmates in skill development, and more broadly, show the import of later-life environments in moderating program efficacy. We view our results as holding general implications for programs beyond those in early childhood. A wide array of programs, such as those targeting health, skill development, job search efforts, may also exhibit similar long-term dependence on participant's surrounding environments. Understanding where these effects arise is essential for optimal program design. We hope that our empirical framework to explore both short-run and long-run impacts along with mechanisms can be useful for these other areas of study.

References

Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin 145*(12), 1103.

Agostinelli, F., C. Avitabile, and M. Bobba (2023). Enhancing human capital in children: A case study on scaling. Technical report, National Bureau of Economic Research.

Al-Ubaydli, O., J. A. List, and D. Suskind (2020). The science of using science: Towards an understanding of the threats to scalability. *International Economic Review 61*(4), 1387–1409.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, Volume 1, pp. 73–140. Elsevier.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., S. Cattan, and C. Meghir (2022). Early childhood development, human capital, and poverty. *Annual Review of Economics 14*(1), 853–892.

Babcock, P. S. and J. L. Hartman (2010). Networks and workouts: Treatment size and status specific peer effects in a randomized field experiment. Technical report, National Bureau of Economic Research.

Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest 21*(2), 55–97.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science 333*(6045), 975–978.

Barrett, G. F. and S. G. Donald (2003). Consistent tests for stochastic dominance. *Econometrica 71*(1), 71–104.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy 70*(5, Part 2), 9–49.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015). The effect of providing peer information on retirement savings decisions. *The Journal of Finance 70*(3), 1161–1201.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy 3*(1), 25–53.

Blair, C. and M. Willoughby (2006a). Measuring executive function in young children: Operation span. *Chapel Hill, NC: Pennsylvania State University and University of North Carolina*.

Blair, C. and M. Willoughby (2006b). Measuring executive function in young children: Spacial conflict ii: Arrows. *Chapel Hill, NC: The Pennsylvania State University and the University of North Carolina at Chapel Hill*.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies 84*(2), 547–578.

Botvin, C. M., J. M. Jenkins, R. C. Carr, K. A. Dodge, D. H. Clements, J. Sarama, and T. W. Watts (2024). Can peers help sustain the positive effects of an early childhood mathematics intervention? *Early Childhood Research Quarterly 67*, 159–169.

Boucher, V., M. Rendall, P. Ushchev, and Y. Zenou (2024). Toward a general theory of peer effects. *Econometrica 92*(2), 543–565.

Bruhn, J. and E. Emick (2023). Lottery evidence on the impact of preschool in the united states: A review and meta-analysis. Technical report, MIT Blueprint Labs.

Burchinal, M., R. Pianta, A. Ansari, J. Whittaker, and V. Vitiello (2023). Kindergarten academic and social skills and exposure to peers with pre-kindergarten experience. *Early Childhood Research Quarterly 62*, 41–52.

Bursztyn, L., G. Egorov, and R. Jensen (2019). Cool to be smart or smart to be cool? understanding peer pressure in education. *The Review of Economic Studies 86*(4), 1487–1526.

Campbell, F. A., E. P. Pungello, S. Miller-Johnson, M. Burchinal, and C. T. Ramey (2001). The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology 37*(2), 231–42.

Cappelen, A., J. List, A. Samek, and B. Tungodden (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy 128*(7), 2739–2758.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics 27*(3), 439–464.

Carrell, S. E., M. Hoekstra, and J. E. West (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy 11*(4), 159–182.

Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics 2*(1), 211–228.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Cascio, E. U. and D. O. Staiger (2012). Knowledge, tests, and fadeout in educational interventions. Technical report, National Bureau of Economic Research.

Castillo, M., J. A. List, R. Petrie, and A. Samek (2020). Detecting drivers of behavior at an early age: Evidence from a longitudinal field experiment. Technical report, National Bureau of Economic Research.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *Quarterly Journal of Economics 128*(2), 531–580.

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

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

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review 104*(7), 2049–2074.

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

Duflo, E. and E. Saez (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics 85*(1), 121–148.

Duncan, G. J., J. Boisjoly, M. Kremer, D. M. Levy, and J. Eccles (2005). Peer effects in drug use and sex among college students. *Journal of Abnormal Child Psychology 33(3)*, 375–385.

Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives 27*(2), 109–132.

Dunn, L. M. and L. M. Dunn (1997). *PPVT-III: Peabody picture vocabulary test.* American Guidance Service.

Durkin, K., M. W. Lipsey, D. C. Farran, and S. E. Wiesen (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology 58*(3), 470–484.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015). Early childhood education. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, pp. 235–297. University of Chicago Press.

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The B.E. Journal of Economic Analysis & Policy 8*(1).

Fletcher, J. M. and M. Tienda (2009). High school classmates and college success. *Sociology of Education 82*(4), 287–314.

Fryer, R. G., S. D. Levitt, J. A. List, and A. Samek (2020). Introducing cogx: A new preschool education program combining parent and child interventions. Technical report, National Bureau of Economic Research.

Gibbs, C., J. Ludwig, and D. L. Miller (2013). Head start origins and impacts. *Legacies of the War on Poverty*, 39–65.

Gormley Jr, W. T., T. Gayer, D. Phillips, and B. Dawson (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology 41*(6), 872–884.

Gray, S. W. and R. A. Klaus (1970). The early training project: A seventh-year report. *Child Development*, 909–924.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2023). The long-term effects of universal preschool in boston. *Quarterly Journal of Economics 138*(1), 363–411.

Guo, K., A. Islam, J. List, M. Vlassopoulos, and Y. Zenou (2024). Early childhood education, parental social networks, and child development.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*, 689–733.

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

Herbst, M., A. Sobotka, and P. Wójcik (2023). The effect of peer group stability on achievements: Evidence from poland. *European Journal of Education 58*(1), 166–180.

Hill, A. J. and D. B. Jones (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review 64*, 1–12.

Holz, J. E., R. G. Rivera, and B. A. Ba (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy 15*(2), 256–291.

Jenkins, J., T. Watts, and K. Dodge (2024). In search of dynamic complementarities between early and later education: Evidence from north carolina's pre-k and k–12 school funding reforms. In *AEA Papers and Proceedings*, Volume 114, pp. 474–479. American Economic Association.

Jenkins, J. M., T. W. Watts, K. Magnuson, E. T. Gershoff, D. H. Clements, J. Sarama, and G. J. Duncan (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *Journal of Research on Educational Effectiveness 11*(3), 339–374.

Joensen, J. S., J. A. List, A. Samek, and H. Uchida (2022). Using a field experiment to understand skill formation during adolescence. *SSRN*.

Joensen, J. S. and H. S. Nielsen (2018). Spillovers in education choice. *Journal of Public Economics 157*, 158–183.

Johnson, R. C. (2024). Synergistic impacts of expansions in pre-k access and school funding on student achievement: Evidence from california's transitional kindergarten rollout. In *AEA Papers and Proceedings*, Volume 114, pp. 467–473. American Economic Association.

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

Larroucau, T., I. Rios, A. Fabre, and C. Neilson (2024). College application mistakes and the design of information policies at scale.

Lavy, V. and E. Sand (2012). The friends factor: How students' social networks affect their academic achievement and well-being? Technical report, National Bureau of Economic Research.

Lavy, V. and A. Schlosser (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Lee, V. E. and S. Loeb (1995). Where do head start attendees end up? one reason why preschool effects fade out. *Educational Evaluation and Policy Analysis 17*(1), 62–82.

Lipsey, M. W., D. C. Farran, and K. Durkin (2018). Effects of the tennessee prekindergarten program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly 45*, 155–176.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale.* Crown Currency.

List, J. A. (2024). Optimally generate policy-based evidence before scaling. *Nature 626*(7999), 491–499.

List, J. A. (2025). *Experimental Economics: Theory and Practice.* University of Chicago Press.

List, J. A., F. Momeni, M. Vlassopoulos, and Y. Zenou (2020). Neighborhood spillover effects of early childhood interventions. Technical report, National Bureau of Economic Research.

Magnuson, K. A., C. Ruhm, and J. Waldfogel (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly 22*(1), 18–38.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–145.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy 66*(4), 281–302.

Mobarak, A. M. (2022). Assessing social aid: the scale-up process needs evidence, too. *Nature 609*(7929), 892–894.

Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics 92*(3), 562–576.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer (2012). Third grade follow-up to the head start impact study: Final report. opre report 2012-45. *Administration for Children & Families*.

Ramey, C. T., F. A. Campbell, M. Burchinal, M. L. Skinner, D. M. Gardner, and S. L. Ramey (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science 4*(1), 2–14.

Rivera, R. (2022). The effect of minority peers on future arrest quantity and quality. Technical report, Technical report.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annual Review of Economics 6*(1), 253–272.

Schaub, S. and M. S. Schaub (2024). Package 'stodom'.

Schultz, T. W. (1961). Investment in human capital. *American Economic Review 51*(1), 1–17.

Schweinhart, L. J. et al. (1993). *Significant Benefits: The High/Scope Perry Preschool Study through Age 27. Monographs of the High/Scope Educational Research Foundation, No. Ten.* ERIC.

Smith-Donald, R., C. C. Raver, T. Hayes, and B. Richardson (2007). Preliminary construct and concurrent validity of the preschool self-regulation assessment (psra) for field-based research. *Early Childhood Research Quarterly 22*(2), 173–187.

U.S. Census Bureau (2010). 2010 census.

Wan, S., T. N. Bond, K. Lang, D. H. Clements, J. Sarama, and D. H. Bailey (2021). Is intervention fadeout a scaling artefact? *Economics of Education Review 82*, 102090.

Wang, S. and D. Y. Yang (2021). Policy experimentation in china: The political economy of policy learning. Technical report, National Bureau of Economic Research.

Weikart, D. P. (1970). Longitudinal results of the ypsilanti perry preschool project.

Woodcock, R. W., K. S. McGrew, N. Mather, et al. (2001). Woodcock-johnson iii tests of achievement.

Yoshikawa, H., C. Weiland, J. Brooks-Gunn, M. R. Burchinal, L. M. Espinosa, W. T. Gormley, J. Ludwig, K. A. Magnuson, D. Phillips, and M. J. Zaslow (2013). Investing in our future: The evidence base on preschool education. *Society for Research in Child Development*.

Appendix

A Literature review figure

Figure 1 shows, in the spirit of Bruhn and Emick (2023), the effects on academic achievements documented from past work evaluating demonstration programs (Perry Preschool Project (Anderson, 2008; Elango et al., 2015), Carolina Abecedarian Project (Anderson, 2008; Elango et al., 2015), Early Training Project (Anderson, 2008; Elango et al., 2015)), Head Start (Deming, 2009; Kline and Walters, 2016), local universal preschool programs (Boston (Gray-Lobe et al., 2023), Tennessee Voluntary Pre-kindergarten Program (Lipsey et al., 2018; Durkin et al., 2022), Oklahoma (Gormley Jr et al., 2005; Fitzpatrick, 2008)), and the preschool studied in this paper, Chicago Heights Early Childhood Center (Castillo et al., 2020).

Anderson (2008) provides program effects separately by gender. We follow Bruhn and Emick (2023) and pool these gender-specific estimates into a precision-weighted average effect. For some of the papers, age was inferred from grade level or years after program implementation. If effects were provided at the grade level (e.g. grade 3 test scores), we converted to age assuming an "on-time" progression (grade 3 at age 8, grade 4 at age 9, and so on). If the standard deviation of the outcome measure was not provided (e.g. for IQ in ages between 5-14 for Anderson (2008)) then we took the closest reported one in terms of age. If an age range was provided, we take the midpoint of the range. For significance level, if the p-value corresponding to the effect size was not presented, then we approximated it using $2\phi(|\frac{\beta}{s}|)$ where $\phi(\cdot)$ is the cdf of the normal distribution, β is the effect size and s is the reported standard error (both in standard deviation units).

B Measures

B.1 Measures from Chicago Heights Early Childhood Center

Children who participated in CHECC were assessed on cognitive and executive function skills, both before and after program implementation.

For cognitive skills, we implemented standard, nationally-normed tests aimed at measuring math, letter and word recognition, writing, and receptive vocabulary. A number of these used the Woodcock-Johnson Test of Achievement (WJ-III) (Woodcock et al., 2001): math ability was measured using the WJ-III applied problems and quantitative concepts subtests, letter and word recognition using the spelling subtest, and writing ability using the spelling subtest. Receptive vocabulary was measured using the Peabody Picture Vocabulary Test III (Dunn and Dunn, 1997).

For executive function skills, we evaluate working memory, inhibitory control, emotion regulation, and attention in two ways: the Preschool Self-Regulation Assessment (PRSA) (Smith-Donald et al., 2007), and the Blair and Willoughby Measures of Executive Function (Blair and Willoughby, 2006a,b).

The PRSA is a survey completed by assessors on a child's behaviors. The assessors were not told of a child's treatment status and were asked to evaluate the child on a point scale according to various statements such as "pays attention during instructions". Because a normed standard for these tests were not available, following (Castillo et al., 2020) who use the same data, we use the percentage correct or points attained as the score in our analysis.

Measures of Executive Function contained various tasks that children answered in paper-and-pencil format. For instance, the "Operation Span" task measures working memory by showing children a sequence of images of animals in various types of houses, and then asks to recall which animals were in which houses.

We use each measure to estimate a latent cognitive skill measure and a latent noncognitive skill measure, using confirmatory skill analysis, as implemented in Castillo et al. (2020), and described in Heckman et al. (2013). We explain these steps in Section C.

B.2 Measures from elementary school

We receive administrative data on standardized test scores, disciplinary referrals, course grades, absenteeism from Chicago Heights school district.

The district implements two types of standardized tests, Measures of Academic Progress (MAP) and the Illinois Assessment of Readiness (IAR) which was formerly the Partnership for Assessment of Readiness for College and Careers (PARCC). MAP exams are conducted three times a year, starting from 2nd grade, in the subjects math and reading. MAP scores are received as an RIT scale from the Northwest Evaluation Association (NWEA), meaning that the scale extends equally across grades. MAP tests are computer adaptive and taken by students during school. IAR/PARCC exams are conducted once a year, starting from 3rd grade.

With regards to course grades, we follow Castillo et al. (2020) and take the grade point average (GPA) from math and English language arts courses. For disciplinary infractions, we consider the indicator for whether a student has any infractions, and the count of infractions.

C Skills and measurement error

C.1 Identifying skills

We now expand on our approach described in Section 2.4. We apply the factor analysis steps explained by Heckman et al. (2013). Using the notation in Section 2.4 and omitting time and individual subscripts for simplicity, we assume that (1) measurement error across the measures are independent, $Cov(\eta^m, \eta^{m\prime}) = 0$ for all $m \neq m\prime$, and (2) measurement errors are uncorrelated with latent skill, $Cov(\eta^m, \theta) = 0$ for all m.

We set the scale and location of skills by first setting it to be mean zero without loss of generality: $\mathbb{E}[\ln \theta] = 0$. We normalize the cognitive skill factor scale to its first measure, meaning that we set $\lambda^1 = 1$. Setting the mean of the skill factor allows us to identify the factor intercept: $\mu^m = \mathbb{E}[Z^m]$.

Heckman et al. (2013) show that factor loadings can be identified from the covariances of observed measures as long as there are at least three dedicated measures, which is the case in this paper. To see it explicitly, given the previous assumptions, we have $\lambda^2 = \frac{Cov(M^2,M^3)}{Cov(M^1,M^3)}$ and $\lambda^3 = \frac{Cov(M^2,M^3)}{Cov(M^1,M^2)}$.

C.2 Estimating skills

We estimate the cognitive skill factor using math test scores, reading test scores, and GPA. The scale is set using math test scores. Each component is standardized before estimating the measurement system. Table A1 summarizes the factor loadings.

Table A1: Cognitive Skill Factor

	Loading	Signal	Noise
Math Test Score (std)	1.000	0.894	0.106
Reading Test Score (std)	0.872***	0.679	0.321
	(0.013)		
GPA (std)	0.644***	0.336	0.664
	(0.018)		
Factor Variance	0.784		

Notes: This table shows the estimated factor loadings, and the signal and noise ratios for each measure. The ratios reflect the proportion of the variance that is from signal and noise, respectively. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$.

D Framework details

$$
\begin{aligned}
\delta &= \mathbb{E}\left[\Big(\underbrace{Y_{D=1,N=1} - Y_{D=1,N=0}}_{\text{effect of preschool with high exposure}} \Big) - \Big(\underbrace{Y_{D=0,N=1} - Y_{D=0,N=0}}_{\text{effect of preschool with low exposure}} \Big)\right] \\
&= \mathbb{E}[Y_{1,1} - Y_{1,0}|D = 1] - \mathbb{E}[Y_{0,1} - Y_{0,0}|D = 0] && \text{by Assump. 1} \\
&= \mathbb{E}[Y_{1,1} - Y_{1,0}|D = 1, \gamma] - \mathbb{E}[Y_{0,1} - Y_{0,0}|D = 0, \gamma] && \text{by Assump. 2} \\
&= \mathbb{E}[Y_{1,1}|D = 1, \gamma] - \mathbb{E}[Y_{1,0}|D = 1, \gamma] - \\
&\qquad \mathbb{E}[Y_{0,1}|D = 0, \gamma] + \mathbb{E}[Y_{0,0}|D = 0, \gamma] \\
&= \mathbb{E}[Y_{1,1}|D = 1, N = 1, \gamma] - \mathbb{E}[Y_{1,0}|D = 1, N = 0, \gamma] \\
&\qquad - \mathbb{E}[Y_{0,1}|D = 0, N = 1, \gamma] + \mathbb{E}[Y_{0,0}|D = 0, N = 0, \gamma] && \text{by Assump. 3} \\
&= \mathbb{E}[Y|D = 1, N = 1, \gamma] - \mathbb{E}[Y|D = 1, N = 0, \gamma] \\
&\qquad - \mathbb{E}[Y|D = 0, N = 1, \gamma] + \mathbb{E}[Y|D = 0, N = 0, \gamma] && \text{(A1)}
\end{aligned}
$$

E Additional Figures and Tables

Table A2: Preschool balance test

Variable	(1) Control	(2) PK	(3) PK v. Control
Age of child at exp./recruit (months)	48.97	45.12	-3.84***
	(12.83)	(6.93)	(0.00)
Male	0.50	0.49	-0.01
	(0.50)	(0.50)	(0.73)
Black	0.24	0.22	-0.02
	(0.43)	(0.42)	(0.59)
Hispanic	0.61	0.56	-0.05
	(0.49)	(0.50)	(0.28)
Birth weight (Lbs)	7.15	7.21	0.06
	(1.29)	(1.22)	(0.64)
Age of mother at child's birth	26.84	25.60	-1.23**
	(6.43)	(5.99)	(0.04)
Mother Education: Did not complete High School (%)	0.32	0.30	-0.02
	(0.47)	(0.46)	(0.64)
Mother Education: High School (%)	0.30	0.36	0.06
	(0.46)	(0.48)	(0.16)
Mother Education: Bachelor (%)	0.05	0.02	-0.03*
	(0.21)	(0.13)	(0.10)
Mother Education: Master's degree and above (%)	0.03	0.03	-0.00
	(0.18)	(0.17)	(0.89)
Two-parent household (%)	0.81	0.78	-0.03
	(0.39)	(0.41)	(0.46)
Income (Yearly, dollars)	21198.54	20947.47	-251.07
	(17173.09)	(14883.01)	(0.87)
Cognitive Skill	-0.13	-0.06	0.07
	(0.63)	(0.62)	(0.21)
Letter and Word Recognition (WJ-III)	33.20	32.11	-1.10
	(26.13)	(26.04)	(0.64)
Writing Skills (WJ-III)	30.80	31.88	1.08
	(25.60)	(27.61)	(0.64)
Math - Applied Problems (WJ-III)	33.51	36.10	2.59
	(24.29)	(24.46)	(0.23)
Math - Quantitative Concepts (WJ-III)	40.81	43.91	3.09
	(20.63)	(22.26)	(0.11)
Receptive Vocabulary (PPVT-III)	25.71	25.90	0.18
	(23.47)	(23.02)	(0.93)
Noncognitive Skill	-0.03	-0.02	0.01
	(0.60)	(0.68)	(0.82)
Attention (PSRA Assessment)	0.73	0.71	-0.02
	(0.21)	(0.22)	(0.31)
Emotion Regulation (PSRA Assessment)	0.34	0.34	-0.00
	(0.31)	(0.29)	(0.94)
Working Memory (Blair & Willoughby)	0.64	0.65	0.01
	(0.25)	(0.27)	(0.53)
Inhibitory Control (Blair & Willoughby)	0.49	0.52	0.03
	(0.29)	(0.30)	(0.46)
Observations	440	203	643

Notes: This table shows the mean pre-CHECC characteristics for children in the Chicago Heights elementary school sample used in this paper, who participated in CHECC and were assigned to (1) Control or (2) Preschool (PK). The third column shows the difference in means between control and Preschool children, with stars corresponding to p-values from a t-test comparison of means. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A3: Effect of Preschool on district and school choice

		If student enrolls in given school								
	CH	S1	S2	S3	S4	S5	S6	S7	S8	S9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Preschool	0.00	0.02	0.01	-0.01	-0.00	-0.00	-0.00	0.00	-0.01	-0.00
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
N	2047	2047	2047	2047	2047	2047	2047	2047	2047	2047

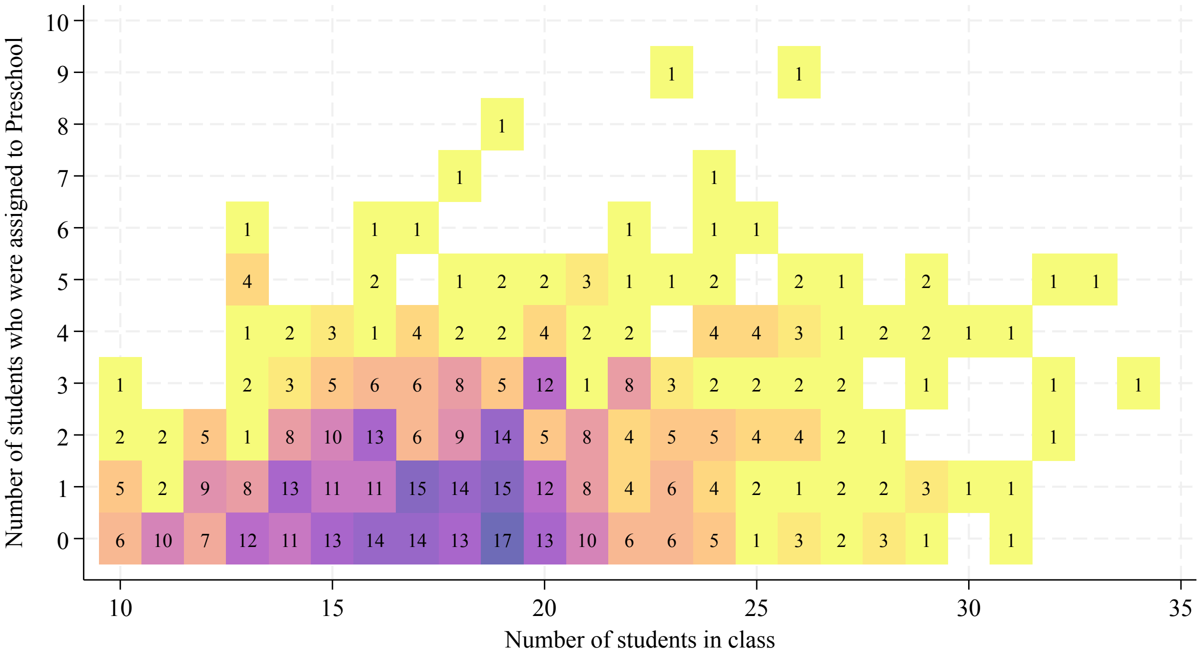
Notes: This table shows the effects of preschool on whether a student ever enrolls in a school in Chicago Heights school district. Dependent variable for first column is an indicator for whether a student ever enrolls in any elementary school in Chicago Heights District during the time range of the study. The subsequent 9 columns are whether a student ever enrolls in one of the nine elementary schools in the district, where each column corresponds to a unique school. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A4: Placebo tests

	$N^{NonCHECC}$	$N^{Control}$	$N^{Preschool}$
Race: White	0.00	0.00	0.00
	(0.00)	(0.00)	(0.01)
Race: Black	0.01	0.00	-0.01*
	(0.00)	(0.00)	(0.00)
Hispanic: Yes	-0.01	0.00	0.01***
	(0.00)	(0.00)	(0.01)
Gender: Female	0.00	0.00	0.00
	(0.00)	(0.00)	(0.01)
Cognitive Skill (year prior)	0.00	-0.01	0.01
	(0.01)	(0.01)	(0.02)
GPA (year prior)	-0.01	0.00	0.02
	(0.01)	(0.01)	(0.01)
MAP math (STD) (year prior)	0.00	-0.01	0.02
	(0.01)	(0.01)	(0.02)
MAP reading (STD) (year prior)	0.02	-0.02	0.01
	(0.01)	(0.01)	(0.02)
Has Discipline (year prior)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Discipline (ihs) (year prior)	0.00	0.00	-0.01
	(0.00)	(0.01)	(0.01)
N Discipline (year prior)	-0.01	0.02	-0.02
	(0.02)	(0.02)	(0.02)
Fraction of School Days Absent (year prior)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)

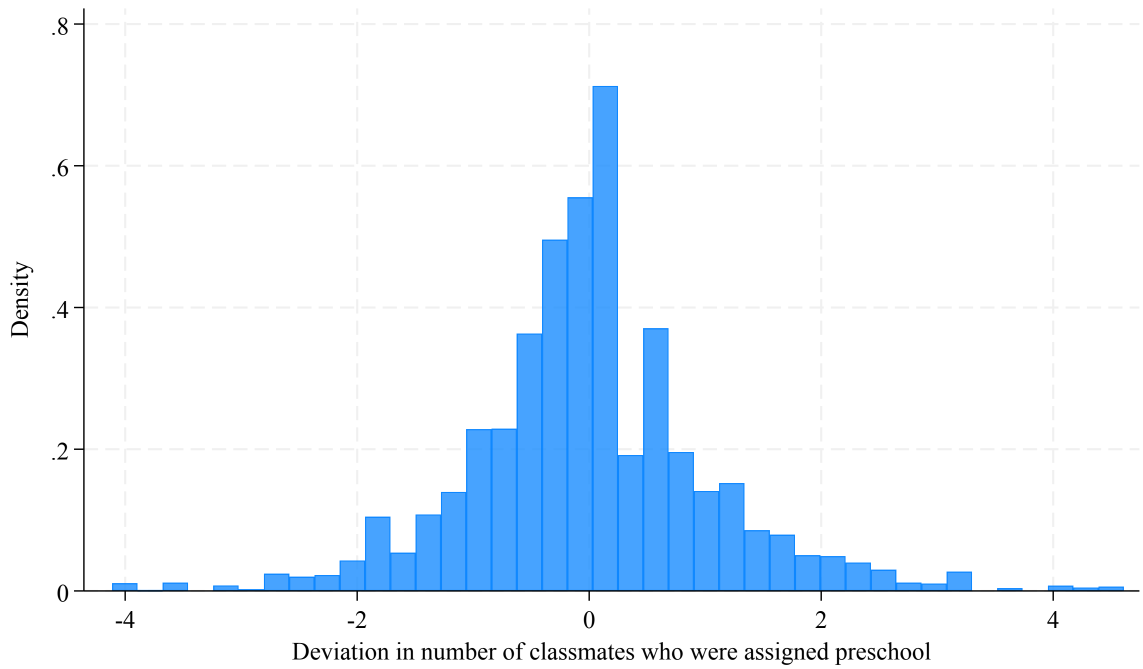
Notes: This table shows the relationship between classmate composition and a pre-determined student characteristic. The sample is all first through 5th graders in our sample, including non-CHECC students. Each cell represents the coefficient from its own regression (π_1 from eq. 8). An observation is at the student-year level. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Figure A1: Joint distribution of classroom sizes



Notes: This figure shows the joint distribution between a classroom's total number of students and number of students who were assigned to preschool. Cell labels correspond to the number of unique classroom observations.

Figure A2: Deviations in number of elementary school classmates who were assigned to Preschool



Notes: This figure shows the distribution of the de-meaned number of a student's number of classmates who were assigned to preschool. De-meaning is done at the school-grade-year level, and also residualizes out the total number of classmates in the student's class. An observation is at the student-year level.

Table A5: Attrition after Kindergarten and first grade

	Classmates during...								
	Kindergarten					1st Grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
N Preschool classmates	0.01	0.02	0.02	0.02	0.03	0.01	0.01	-0.01	0.00
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
N Preschool classmates × Preschool	-0.02	-0.00	0.01	-0.03	-0.02	0.01	-0.01	-0.00	-0.00
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Grade in Data	1	2	3	4	5	2	3	4	5
School × Grade × Year FE	×	×	×	×	×	×	×	×	×
Demographics	×	×	×	×	×	×	×	×	×
N	1870	1870	1870	1870	1870	1844	1844	1844	1844
Adjusted R^2	0.13	0.21	0.21	0.22	0.23	0.37	0.30	0.21	0.22

Notes: This table shows the effect of classmate composition on the likelihood that the student remains in the sample in subsequent grades 2-5, corresponding to the outcome in Table 2. Dependent variable is an indicator for whether the student is in the sample for a given grade: (1) first (2) second (3) third (4) fourth (5) fifth (6) second (7) third (8) fourth (9) fifth. The regression specification is the same as in columns 2 and 4 of Table 2, except for the dependent variable. Heteroskedastic robust standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A6: Effects of classmate composition on cognitive skill components

	Outcome: Math test score					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.04	-0.05	0.01	-0.02	0.01	-0.01
	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)
N Preschool classmates × Preschool		0.03		0.08*		0.08**
		(0.05)		(0.04)		(0.03)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4295	4295	4681	4681	5695	5695
N Students	1519	1519	1659	1659	2242	2242
Adjusted R^2	0.21	0.21	0.21	0.21	0.19	0.19
	Outcome: Reading test score					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.04*	-0.06*	0.03	-0.01	0.03**	0.00
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
N Preschool classmates × Preschool		0.05		0.08*		0.08**
		(0.04)		(0.04)		(0.03)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4305	4305	4688	4688	5694	5694
N Students	1525	1525	1665	1665	2242	2242
Adjusted R^2	0.24	0.24	0.23	0.23	0.18	0.18
	Outcome: GPA out of 4					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.02	-0.02	0.01	-0.00	0.00	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)
N Preschool classmates × Preschool		0.00		0.08**		0.08***
		(0.04)		(0.04)		(0.03)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4941	4941	5382	5382	6825	6825
N Students	1523	1523	1659	1659	2409	2409
Adjusted R^2	0.15	0.15	0.17	0.17	0.24	0.24

Notes: This table shows the effect of having elementary school classmates during (columns 1-2) Kindergarten or (columns 3-4) 1st grade or (columns 5-6) grades 2-5 who were randomly assigned to Preschool, on the cognitive skill component measures during grades 2-5. Dependent variable is given by the column heading. Sample contains student-year observations during grades 2-5. Each column follows Equation 9, except for columns 1-4 which do not include fixed effects for grades given that they focus on classmate composition from a single grade. "N Preschool classmates" is the number of classmates during the corresponding grade who were randomly assigned to Preschool. School-by-grade-by-year fixed effects correspond to the school and year of the grade indicated in the column header. Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to Preschool. For additional precision, we include measures collected from CHECC (age at experiment, birth weight, mother's education, mother's age at child's birth, family income, indicator for two-parent household, pre-elementary school cognitive and executive function skills), and indicators for each on whether the student is missing an observation for the indicator. Students who were not part of CHECC are imputed as the sample mean value. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A7: Effects of classmate composition on national test score percentile ranks

	Outcome: Math national percentile rank					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-1.14*	-1.30	0.13	-0.73	0.12	-0.63
	(0.65)	(0.85)	(0.65)	(0.82)	(0.39)	(0.57)
N Preschool classmates × Preschool		0.75		1.90*		1.95**
		(1.17)		(1.05)		(0.85)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4240	4240	4624	4624	5617	5617
N Students	1468	1468	1606	1606	2170	2170
Adjusted R^2	0.20	0.21	0.20	0.20	0.18	0.19
	Outcome: Reading national percentile rank					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-1.07	-1.48*	1.04	-0.14	0.73	-0.33
	(0.72)	(0.89)	(0.72)	(0.90)	(0.45)	(0.66)
N Preschool classmates × Preschool		1.49		2.32*		2.86***
		(1.26)		(1.23)		(0.98)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4246	4246	4627	4627	5602	5602
N Students	1470	1470	1608	1608	2161	2161
Adjusted R^2	0.22	0.23	0.21	0.21	0.16	0.17

Notes: This table shows the effect of having elementary school classmates during (columns 1-2) Kindergarten or (columns 3-4) 1st grade or (columns 5-6) grades 2-5 who were randomly assigned to Preschool, on the national test score (MAP) percentile ranks during grades 2-5. Dependent variable is given by the column heading. Sample contains student-year observations during grades 2-5. Each column follows Equation 9, except for columns 1-4 which do not include fixed effects for grades given that they focus on classmate composition from a single grade. "N Preschool classmates" is the number of classmates during the corresponding grade who were randomly assigned to Preschool. School-by-grade-by-year fixed effects correspond to the school and year of the grade indicated in the column header. Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to Preschool. For additional precision, we include measures collected from CHECC (age at experiment, birth weight, mother's education, mother's age at child's birth, family income, indicator for two-parent household, pre-elementary school cognitive and executive function skills), and indicators for each on whether the student is missing an observation for the indicator. Students who were not part of CHECC are imputed as the sample mean value. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A8: Effects of classmate composition on MAP standardized test RIT scores

	Outcome: Math RIT score					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.72*	-0.97*	0.58	0.03	0.27	-0.20
	(0.40)	(0.50)	(0.38)	(0.47)	(0.23)	(0.35)
N Preschool classmates × Preschool		0.70		1.04		1.36***
		(0.66)		(0.64)		(0.51)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4246	4246	4627	4627	5605	5605
N Students	1470	1470	1608	1608	2164	2164
Adjusted R^2	0.44	0.44	0.42	0.43	0.39	0.39
	Outcome: Reading RIT score					
	Kindergarten		1st Grade		Grade 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
N Preschool classmates	-0.58*	-0.59	0.04	-0.48	0.07	-0.42
	(0.33)	(0.43)	(0.32)	(0.40)	(0.20)	(0.30)
N Preschool classmates × Preschool		0.33		1.14**		1.46***
		(0.57)		(0.53)		(0.43)
School × Grade × Year FE	×	×	×	×	×	×
Demographics	×	×	×	×	×	×
N	4240	4240	4624	4624	5617	5617
N Students	1468	1468	1606	1606	2170	2170
Adjusted R^2	0.51	0.52	0.52	0.52	0.50	0.50

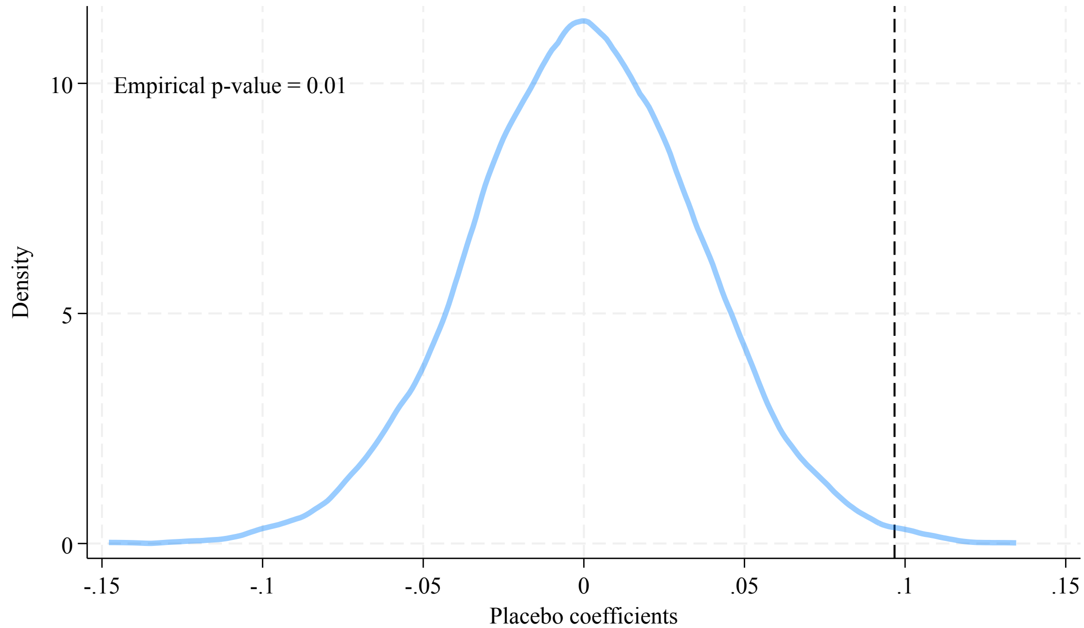
Notes: This table shows the effect of having elementary school classmates during (columns 1-2) Kindergarten or (columns 3-4) 1st grade or (columns 5-6) grades 2-5 who were randomly assigned to Preschool, on MAP RIT scores during grades 2-5. Dependent variable is given by the column heading. Sample contains student-year observations during grades 2-5. Each column follows Equation 9, except for columns 1-4 which do not include fixed effects for grades given that they focus on classmate composition from a single grade. "N Preschool classmates" is the number of classmates during the corresponding grade who were randomly assigned to Preschool. School-by-grade-by-year fixed effects correspond to the school and year of the grade indicated in the column header. Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to Preschool. For additional precision, we include measures collected from CHECC (age at experiment, birth weight, mother's education, mother's age at child's birth, family income, indicator for two-parent household, pre-elementary school cognitive and executive function skills), and indicators for each on whether the student is missing an observation for the indicator. Students who were not part of CHECC are imputed as the sample mean value. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A9: Heterogeneity

	(1)	(2)	(3)	(4)	(5)
N Preschool classmates × Preschool × Mother edu more than hs	0.00				
	(0.06)				
N Preschool classmates × Preschool × Mother edu hs or less	0.11***				
	(0.04)				
N Preschool classmates × Preschool × Male		0.11**			
		(0.04)			
N Preschool classmates × Preschool × Female		0.08			
		(0.05)			
N Preschool classmates × Preschool × Hispanic			0.11***		
			(0.04)		
N Preschool classmates × Preschool × Not Hispanic			0.01		
			(0.08)		
N Preschool classmates × Preschool × White				0.08**	
				(0.04)	
N Preschool classmates × Preschool × Black				-0.02	
				(0.09)	
N Preschool classmates × Preschool × High skill before Preschool					0.12**
					(0.05)
N Preschool classmates × Preschool × Low skill before Preschool					0.07
					(0.05)
School × Grade × Year FE	×	×	×	×	×
Demographics	×	×	×	×	×
p-value for difference in coefficients	0.13	0.61	0.25	0.27	0.47
N	5688	5688	5688	5688	5688
N Students	2237	2237	2237	2237	2237
Adjusted R^2	0.21	0.20	0.20	0.20	0.20

Notes: Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to preschool. Each regression also contains the regressor N Preschool classmates × Preschool × Missing Covariate, to account for students for whom we do not observe the covariate. The p-value for difference in coefficients corresponds to the p-value from testing the null hypothesis that the two coefficients in the corresponding column are equal. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Figure A3: Randomization-based inference



Notes: This figure shows the distribution of estimated coefficients for the interaction term N Preschool classmates $\times$ Preschool from equation 9, using randomization-based inference with 5,000 iterations as described in Section 4.3. The test is under the sharp null hypothesis that preschool classmates do not have differential effects on preschool students. At every iteration, the outcome variable (cognitive skill) is shuffled across students within the unit of randomization: school-by-grade-by-year. The empirical p-value corresponds to the proportion of draws with an estimate as extreme in magnitude as the main estimate. The dashed line corresponds to the main estimate presented in Table 2.

Table A10: Alternative specifications of Preschool classmates

	(1)	(2)
Proportion classmates in Preschool	-0.74	
	(0.55)	
Proportion classmates in Preschool × Preschool	2.61***	
	(0.94)	
At least one classmate in Preschool		-0.13
		(0.09)
At least one classmate in Preschool × Preschool		0.31**
		(0.16)
School × Grade × Year FE	×	×
Demographics	×	×
Sample mean of preschool classmate variable	0.09	0.66
N	5688	5688
N Students	2237	2237
Adjusted R^2	0.18	0.20

Notes: This table shows the effect of classmate characteristics on cognitive skills. Sample contains student-year observations during grades 2-5. Demographics include gender, race, ethnicity, and CHECC controls as described in Section 3.2 and an indicator for being assigned to preschool. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A11: Classmate behaviors and Preschool synergy in cognitive skill development

	(1)	(2)	(3)
N Preschool classmates × Preschool	0.10***	0.10***	0.10***
	(0.04)	(0.04)	(0.04)
Average classmate has discipline (prior year)	1.07***		
	(0.18)		
Average classmate N discipline (prior year)		0.05	
		(0.04)	
Average classmate frac. absent (prior year)			1.44
			(1.19)
School × Grade × Year FE	×	×	×
Demographics	×	×	×
N	5688	5688	5652
N Students	2237	2237	2201
Adjusted R^2	0.20	0.20	0.20

Notes: This table shows the effect of classmate characteristics on cognitive skills, controlling for classmate misbehaviors. Sample contains student-year observations during grades 2-5. Demographics include gender, race, ethnicity. Standard errors in parentheses, clustered at the student-level. * $p < .1$, ** $p < .05$, *** $p < .01$