ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 291/26

# Estimating, Bounding, and Decomposing the Effects of a Parent-Directed School Readiness Program

Flávio Cunha, Qinyou Hu, Andrea Salvati, Kenneth I. Wolpin, Rui Zeng

www.rfberlin.com NOVEMBER 2026

# Estimating, Bounding, and Decomposing the Effects of a Parent-Directed School Readiness Program


## Authors

Flávio Cunha, Qinyou Hu, Andrea Salvati, Kenneth I. Wolpin, Rui Zeng



## Reference

**JEL Codes:** C93, I21, I24, I28, J13, J24

**Keywords:** Early Childhood Development, Human Capital Development

**Recommended Citation:** Flávio Cunha, Qinyou Hu, Andrea Salvati, Kenneth I. Wolpin, Rui Zeng (2026): Estimating, Bounding, and Decomposing the Effects of a Parent-Directed School Readiness Program. RFBerlin Discussion Paper No. 291/26



## Access

Papers can be downloaded free of charge from the RFBerlin website: https://www.rfberlin.com/discussion-papers

Discussion Papers of RFBerlin are indexed on RePEc: https://ideas.repec.org/s/crm/wpaper.html

## Disclaimer

*Opinions and views expressed in this paper are those of the author(s) and not those of RFBerlin. Research disseminated in this discussion paper series may include views on policy, but RFBerlin takes no institutional policy positions. RFBerlin is an independent research institute.*

*RFBerlin Discussion Papers often represent preliminary or incomplete work and have not been peer-reviewed. Citation and use of research disseminated in this series should take into account the provisional nature of the work. Discussion papers are shared to encourage feedback and foster academic discussion.*

*All materials were provided by the authors, who are responsible for proper attribution and rights clearance. While every effort has been made to ensure proper attribution and accuracy, should any issues arise regarding authorship, citation, or rights, please contact RFBerlin to request a correction.*

*These materials may not be used for the development or training of artificial intelligence systems.*



**Imprint**

**RFBerlin**
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

# Estimating, Bounding, and Decomposing the Effects of a Parent-Directed School Readiness Program*

Flávio Cunha† Qinyou Hu‡ Andrea Salvati§ Kenneth I. Wolpin¶ Rui Zeng||

October 28, 2026

**Abstract**

We evaluate JumpStart, a parent-directed school-readiness program run by a Houston-area school district, in a randomized trial spanning three cohorts. Being offered the program raises scores on a curriculum-aligned assessment by 7.6 percentage points and parental investment by 0.26 standard deviations. A model of skill formation with endogenous inputs and program participation yields an average treatment effect of completing the program of about 9 percentage points, and attributes roughly a quarter of this effect to higher measured investment. Nonparametric bounds corroborate the average effect, and a mediation analysis that relaxes part of the model's structure supports the decomposition.

Keywords: *Early Childhood Development, Human Capital Development*
JEL Codes: C93, I21, I24, I28, J13, J24

*We thank the Family and Community Engagement Department of the Alief Independent School District for its support of the study, especially Shanceler Terry and Laura Escamilla. We also thank the Alief family liaisons for their help in recruiting families for the study and Yanitza Velez, our project coordinator. A grant by Arnold Ventures (formerly the Laura and John Arnold Foundation) supported this research. Cunha also acknowledges support from the National Institutes of Health, grant 1R01HD073221-01A1. An earlier version of this paper circulated under the title *"An Evaluation of the Alief Independent School District JumpStart Program: Using a Model to Recover Mechanisms from an RCT."*

†Rice University and NBER; Email: Flavio.Cunha@rice.edu.
‡The University of Texas at San Antonio; Email: qinyou.hu@utsa.edu.
§Aarhus University and TrygFonden's Centre for Child Research; Email: andrea.salvati@econ.au.dk.
¶University of Pennsylvania.
||Central University of Finance and Economics; Email: rui.zeng@cufe.edu.cn.

# 1 Introduction

Children from low-income families begin kindergarten roughly one standard deviation behind their higher-income peers in reading and mathematics achievement (Reardon and Portilla, 2016), and these gaps persist as children progress through school (Bond and Lang, 2013). A substantial body of evidence traces part of this disparity to the early home environment, where disadvantaged children face deficits in the time devoted to learning activities (Kalil et al., 2012), in access to learning materials (Bassok et al., 2016), and in the routines that generate rich interactions with adults (Putnam, 2016). Because foundational skills are most malleable in early childhood, and because later investments build on the skills that early investments create (Cunha and Heckman, 2007), interventions that reach children through their parents before school begins are a natural policy response.

Whether parent-directed programs can deliver on that promise appears to depend on how they are implemented. Small-scale home-visiting trials with trained and closely supervised staff, such as the Jamaican intervention (Gertler et al., 2014, 2021) and the Nurse Family Partnership (Heckman et al., 2017), have produced large and lasting effects on child development. Evidence from programs operated at scale is less encouraging, particularly in the United States. The home-visiting component of Early Head Start achieved only modest impacts (Love et al., 2005), Even Start produced no significant effects on parental investment or child outcomes at a cost of about $9,400 per family per year (St. Pierre et al., 2003), and the positive results from one small trial of HIPPY failed to replicate in the following year's cohort (Baker et al., 1998, 1999).[1] Recent adaptations of the Jamaican home-visiting model show that such programs can remain effective at scale (Meghir et al., 2023; Zhou et al., 2023). In each case, however, the program itself is a transplant: an externally designed protocol, adapted to a new setting, whose content arrives already validated by the original trial. What remains largely untested is the arrangement school districts can actually afford and are most likely to choose: a program the district itself designs, staffs, and operates.

This paper evaluates such a program. JumpStart, a program of the Houston-area Alief Independent School District (AISD), is a parent-directed intervention that prepares three-year-old children for entry into the district's pre-K classrooms. Its curriculum was constructed by asking the district's own pre-K teachers what entering children should know, and it is delivered by district personnel with district resources. Over seven months, parents meet three times a month with an AISD family liaison at their local elementary school, and once more at the end of the program

[1] The Even Start figure is total resources per family in 2000–01, of which half were federal Even Start funds, $4,708 per family, and half local matching funds and in-kind contributions (St. Pierre et al., 2003, Exhibits 1.3 and 2.40). It is about $17,000 in October 2024 dollars.

year. Children attend one of these meetings each month, and families receive learning materials for the home. The program costs roughly $450 per family. JumpStart is thus what a scaled program looks like when the scaling institution is the school district itself. We ask two questions. Does a program of this kind improve the home learning environment and the school readiness of economically disadvantaged children? And if it does, how much of the effect runs through the parental investments the program is built to change, as opposed to the program's own contact with the household?

We conducted a randomized evaluation across three annual cohorts, drawing families from the catchment areas of all 24 AISD elementary schools. Access was assigned by lottery among registered families, and no family assigned to the control group could participate. The design identifies the effect of the offer directly and, granting that the offer matters only through completion, the effect on the families who finish the program. Control-group children learn rapidly on their own, gaining more than 18 percentage points on the district's curriculum-aligned assessment over the program period. The offer of JumpStart added 7.6 percentage points to that gain, and with 78 percent of offered families completing the program, the effect on the families who complete it is 9.8 percentage points. The offer also raised parental investment at home, most visibly in how often parents read to the child and in the number of children's books at home. On the Bracken School Readiness Assessment, a nationally normed test whose content largely lies outside the JumpStart curriculum, the estimated effect is about 2 percentage points, though not robust to corrections for attrition. The program moved what it taught, and detectably less beyond it.

The second question is harder, because the experiment alone cannot answer it. The offer moved knowledge and investment together, so randomization cannot separate the effect of the program's contact with the household from the effect of the investments that contact induced. A naive mediation regression attributes only 8.7 percent of the impact to measured investment, but its estimate of the return to investment is biased in two ways. Measurement error in the investment factor attenuates it, and investment is endogenous if parents choose how much to invest in response to determinants of readiness we do not observe. We address both biases in two ways, first within a structural model of skill formation with endogenous investment and program completion, and then in a linear instrumental-variables mediation analysis that dispenses with the model's selection and distributional assumptions. Both rest on the exclusion that baseline investment affects endline knowledge only through endline investment, and both attribute about a quarter of the program's effect to measured parental investment. Correcting for measurement error alone brings the naive share close to this figure, suggesting that attenuation accounts for most of the gap. The direct effect of assignment remains large, so the hypothesis that investment mediates the entire

effect is rejected. At the point estimates, most of JumpStart's effect does not pass through the investments we measure, although the data cannot rule out an investment channel as large as the direct one.

This division of the effect between the two channels is a fact about JumpStart, not about parenting programs in general. Attanasio et al. (2020) correct the same biases in a Colombian adaptation of the Jamaican model, instrumenting investment with local prices rather than lagged investment, and reach the opposite conclusion. They attribute essentially the entire cognitive impact to stimulation at home, with no direct effect remaining. The comparison carries two lessons. The first is that correcting for these biases does not predetermine the answer, since it attributes the entire effect to investment in one program and a quarter of it in the other. The second is that the share of a program's effect that runs through parental investment does not carry over from one program or population to another. A policy maker who funded parenting programs by their measured effect on home investment alone would therefore overlook much of what a program like JumpStart delivers.

The paper's second contribution is the way its evidence is organized. Every claim is labeled with what supports it. Random assignment identifies the effect of the offer. Adding one-sided compliance and the exclusion that the offer matters only through completion, the ratio of the offer effect to the completion rate identifies the effect on families who would complete the program when offered (Bloom, 1984). The claim that the average effect across all offered families lies between 7.1 and 14.0 percentage points adds only the assumption that completing the program never lowers a child's score (monotone treatment response, Manski, 1997), whose testable implication is not rejected by the experiment. The exact average effect, and the effect on the families who decline the program, rest on the parametric structure of a selection model. The experiment randomizes the offer but not completion, the one cost shifter in our data moves completion only weakly, and whether families sort into completion on unobserved gains is therefore a question our data cannot answer. Rather than resolving that question by assumption and moving on, we measure what the assumption is doing, using the partial-identification framework of Mogstad et al. (2018) to locate where the model's extrapolations are constrained by the data and where they rest on functional form. The same accounting is then carried through the mechanism decomposition, which we show survives without the model's parametric structure. The individual components of this approach are familiar; what this paper offers is an evaluation conducted so that every finding, from the headline effect to the mechanism split, can be traced to the experiment, to a stated assumption, or to a named parametric choice.

We claim no methodological novelty in any of this. The equivalence between instrumental-

variable assumptions and latent-index selection models is Vytlacil's (2002), and the marginal treatment effect framework that unifies the parameters we report is Heckman and Vytlacil's (2005). The bounds are those of Manski (1997) and Manski and Pepper (2000); the framework that confronts the structural model with them is that of Mogstad et al. (2018). The practice of validating structural models against experimental variation goes back at least to Todd and Wolpin (2006), and our mediation analysis borrows the logic of Kline and Walters (2019), who show that parametric selection estimators can be numerically equivalent to instrumental-variable estimands. The sensitivity analysis that prices the mediation's exclusion restriction follows Imai et al. (2010) and Conley et al. (2012), who replace an untestable exclusion with a parameter and re-estimate at every value of it. Kowalski (2023) uses an experiment's own compliance variation to bound effects for those who decline treatment. The case for making structural estimates transparent about their sources of identification has been put by Andrews et al. (2020). What we take from this literature is a discipline rather than a technique, the principle Manski (2003) states as the law of decreasing credibility. Stronger assumptions buy stronger conclusions, and the exchange rate should be visible. This paper applies that discipline to a single evaluation from end to end, so that the reader can see, for each conclusion, what comes from the data and what comes from the model.

The rest of the paper proceeds as follows. Section 2 describes the program; Section 3, the experimental design and its data; and Section 4, the experimental results. Section 5 develops and estimates the model that separates the program's channels, Section 6 establishes what each conclusion owes to the experiment and what it owes to the model, and Section 7 concludes.

## 2 The JumpStart Program

JumpStart was created and designed by staff of the Alief Independent School District (AISD), in southwest Houston, Texas. In the 2017–18 academic year, AISD served over 46,000 students, of whom 83% were economically disadvantaged and 43% were English Language Learners. Since 2015, our team has been working with the Department of Family and Community Engagement (FACE) in the design, implementation, evaluation, and scaling of parent-directed early childhood programs.

To be eligible for JumpStart, parents must live within the AISD catchment area and must have a child who is between 36 and 47 months old on September 1 of each academic year. The program targets low-income families. Among the families who applied at the district's 24 elementary schools, 72% qualify for free or reduced-price lunch and nearly half (48%) receive food stamps, 58% of the parents are Hispanic and 28% are Black, and 34% have completed less than twelve

years of schooling.

The district developed the twenty-two-session curriculum by consulting its own pre-K teachers about the foundational skills children should have upon entering pre-K; the curriculum targets those skills directly. The session-by-session JumpStart curriculum, which includes learning colors, acquiring fine motor skills, counting, name recognition, and book handling, is summarized in Appendix A. Parents meet in groups for one hour with a family liaison at their local elementary school, three times a month over seven months and once more at the end of the program year. The children are present at the third meeting of each month, a frequency which not only provides a chance for the child to interact directly with the liaison, but also enables the family liaison to assess parent fidelity to the curriculum. Thus, JumpStart guides parents to teach foundational skills for the AISD pre-K program and the family liaisons model behavior so parents can use the program materials with their children at home. The cost of the program, accounting for the family liaison's pay, the cost of materials, and the opportunity cost of the parents, is a little under $450 per family per year.[2]

School districts and non-profit organizations in the Houston area already offer parenting programs of this kind, at several times the cost of JumpStart. The Houston Independent School District runs the Home Instruction for Parents of Preschool Youngsters (HIPPY), in which home visitors from the local community bring books and structured activities to families with children aged 2 to 5 every week or two (Westheimer, 2003). Collaborative for Children, a local non-profit, delivers Parents as Teachers (PAT), which offers home visits on child development and developmental screenings from pregnancy to age 4.[3] HIPPY costs $1,548 to $2,322 per family per year and PAT about $3,400 per family in its first year, on average across sites (Corso et al., 2022).[4] Both programs rest on individual home visits, whereas JumpStart meets parents in groups at their local elementary school, which spreads each liaison's time over about ten families.

Similar programs have also been implemented in developing countries. The Jamaican Early Childhood Development Intervention of the 1980s, a pioneering home-visiting program, enrolled stunted children aged 9 to 24 months, whose mothers received weekly visits from community

[2] Appendix A.2 details the calculation.

[3] Evidence on the effects of these programs is thin. The one randomized evaluation of HIPPY found gains that did not replicate in the following cohort (Baker et al., 1998), and two randomized trials of PAT in California found small and inconsistent effects on parents' knowledge, attitudes, and behavior and no gains in child development (Wagner and Clayton, 1999).

[4] The costs of the other programs in this section are in October 2024 dollars, converted with the Bureau of Labor Statistics CPI Inflation Calculator. The HIPPY range updates the $1,000 to $1,500 per family per year, in August 2006 dollars, reported at https://www.idra.org/families-and-communities/hippy/. The JumpStart figure is in dollars of the program years, 2016 to 2019, and is about $570 in October 2024 dollars, still a fraction of the cost of the other programs.

health aides over two years teaching parenting skills and interactive play (Grantham-McGregor et al., 1991). Its long-term gains in cognitive development, educational attainment, and adult earnings (Gertler et al., 2014, 2021) led to replications in Colombia (Attanasio et al., 2014), India (Andrew et al., 2020), and China (Zhou et al., 2023), which enrolled children aged 12 to 24, 10 to 20, and 6 to 25 months respectively, the last at a cost of about $703 per child per year (Zhou et al., 2023). JumpStart differs from these programs in when it begins, since it enrolls children aged 36 to 47 months, around the age at which the Jamaican model and its replications end. While there is broad agreement that intervening early is valuable, there is little consensus on how the returns to parenting interventions vary within early childhood, and evidence on children in the year before formal schooling is scarce. JumpStart thus provides evidence on a later margin, immediately before pre-K entry.

# 3 Experimental Design

## 3.1 Recruitment and Randomization

Recruitment took place at each school through the family liaison. Families seeking a place in JumpStart were told the program was oversubscribed and that spots would be allocated by lottery; any family wanting the program had to enter the lottery, and we are not aware of any family declining for that reason. At enrollment, the accompanying adult, almost always the mother, completed the program application and the child took the baseline JumpStart test. The family was then invited, separately, to join the Center for Economic Mobility evaluation study, with consent obtained under IRB regulations and incentives of twenty-five dollars for the baseline and fifty dollars for the endline assessments; study participation involved a parent survey and the Bracken School Readiness Assessment. Critically, every family that entered the lottery was randomized, whether or not it consented to the study.

The randomization used a stratified block design, identical in all three years. Within each school, families were divided into four blocks by whether the child's baseline JumpStart score and age in months were above or below the lottery-sample means, and treatment was randomly assigned within school and block, with one-half of each school's lottery families assigned to treatment (the extra family going to treatment when the count was odd).

Table 1 summarizes the experiment. Across the three cohorts, which ran in the 2016–17, 2017–18 and 2018–19 school years, 1,216 families entered the lottery: 461 in 24 schools in year one, 389 in 24 schools in year two, and 366 in year three, when two liaison departures reduced coverage to

Table 1: The Experiment at a Glance

| | Cohort 1 | Cohort 2 | Cohort 3 | Pooled |
|---|---|---|---|---|
| Families in the lottery | | | | |
| *Treatment group* | 232 | 199 | 190 | 621 |
| *Control group* | 229 | 190 | 176 | 595 |
| In study at baseline (% of lottery) | | | | |
| *Treatment group* | 90.9 | 92.5 | 94.2 | 92.4 |
| *Control group* | 89.5 | 93.7 | 96.0 | 92.8 |
| In study at endline (% of lottery) | | | | |
| *Treatment group* | 73.3 | 80.4 | 82.6 | 78.4 |
| *Control group* | 55.5 | 72.6 | 78.4 | 67.7 |
| Graduated (% of treatment) | 70.6 | 84.4 | 80.3 | 78.2 |
| Baseline JumpStart score, control mean (%) | 50.9 | 53.4 | 48.4 | 50.9 |

*Note:* All families in the lottery were randomized, whether or not they consented to the study. *Baseline* is the share of lottery families completing the baseline assessment; *Endline* is the share completing the endline assessment, by arm. *Graduated* is the share of treatment-group families in the experiment sample, those observed at endline, completing 16 or more of the 22 sessions (Section 3.4). *Baseline JumpStart score* is the control-group mean JumpStart test score in the baseline in-study sample, in percent of items answered correctly. Appendix Table B1 reports the complete recruitment counts by cohort and Appendix Table B3 the balance of every baseline covariate by sample.

22 schools. The year-one recruitment strategy, a target of twenty families per school with individual liaison incentives, filled the lottery but produced weak attachment to the program among some recruits; in response, years two and three used a lower individual target with a group incentive for additional families, and attachment improved (Section 3.4). Study participation was high throughout, 92.6 percent of lottery families pooled, and endline response rose across cohorts. Endline response was lower in the control group, with the differential concentrated in year one; Section 4.5 and Appendix C.6 show the results are robust to this differential attrition.

### 3.2 Measurements

Both assessment rounds administered two tests. The JumpStart test, designed by AISD, has 26 items covering the program's curriculum: colors, fine motor skills, counting, sorting, name recognition, and book handling. The Bracken School Readiness Assessment, Third Edition (Bracken, 2007), is a nationally normed test of five concepts, colors, letters, numbers and counting, sizes and comparisons, and shapes, with 85 items in total; only its colors module coincides with the Jump-

Start test, item for item, so it measures readiness beyond the taught content. We score both tests as the percentage of items answered correctly, and from the Bracken we additionally construct an age-adjusted percentile and an indicator for children classified as delayed or very delayed under the publisher's criteria. Appendix B.3 details both instruments, the administration protocol, and the derived measures.

The baseline and endline surveys measured parental involvement through five items, all reported by the parent: the time spent reading to the child, help with school-readiness skills (the alphabet, colors, numbers, and shapes), the number of children's books in the home, the hours per day the television is on while the child is at home, and the frequency of grocery-store trips with the child. Appendix B.3 gives the exact questions and codings. These items are the inputs to the parental-investment factor scores estimated as described in Appendix B.4.

The surveys also collected the participating guardian's age, schooling, race and ethnicity, primary language, and employment status; the child's age in months, sex, and primary language; and household measures including the number of children, food-stamp receipt, and an index of economic well-being whose construction is detailed in Appendix B.5. Appendix Table B2 reports the mean and standard deviation of the test scores and parental inputs at baseline and endline, and of the demographic characteristics at baseline, in the experiment sample.

### 3.3 Balance and Baseline Levels

Randomization delivered balanced arms. Across sixteen baseline characteristics, including the two blocking variables, no treatment–control difference is significant at the 5 percent level, in either the full sample of baseline consenters or the experiment sample of families observed at both rounds; Appendix Table B3 reports every comparison by sample, and Appendix B.2 discusses the largest difference, in home-to-school distance, the only one with a $p$-value below 0.10, at 0.09. The baseline levels shown in Table 2 motivate the intervention. The pooled mean JumpStart score is 51.0 percent, stable across cohorts, with wide dispersion: about a fifth of children score below 25 percent and another fifth above 75 percent. The Bracken baseline is far lower, 23.9 percent on average, and 50.7 percent of the children in the study are classified as delayed or very delayed in school readiness under the publisher's criteria. The program is thus serving a population in which half the children start behind the readiness threshold, which is the margin the district designed JumpStart to move.

Table 2: Baseline Test Scores Distribution in the Experiment Sample

| | Cohort 1 | Cohort 2 | Cohort 3 | Pooled |
|---|---|---|---|---|
| JumpStart test score | | | | |
| *Mean* | 50.14 | 53.92 | 48.77 | 50.96 |
| *Std. deviation* | 26.41 | 24.97 | 25.56 | 25.72 |
| *% lower than 25%* | 20.88 | 16.44 | 21.69 | 19.66 |
| *% higher than 75%* | 20.54 | 22.82 | 17.63 | 20.34 |
| Bracken test score | | | | |
| *Mean* | 23.62 | 25.05 | 22.90 | 23.86 |
| *Std. deviation* | 19.09 | 20.17 | 17.71 | 19.02 |
| *% delayed or very delayed* | 52.36 | 48.66 | 51.19 | 50.73 |

*Note:* Baseline JumpStart and Bracken scores, in percent of items answered correctly, for the experiment sample, by cohort and pooled. *% lower than 25%* and *% higher than 75%* are the shares of children in the tails of the JumpStart score distribution. *% delayed or very delayed* applies the age-specific SRC raw-score cut-offs of Bracken (2007).

### 3.4 Program Attendance

Families offered the program attended 17.6 of the 22 sessions on average, and 78.2 percent graduated, which requires attending at least 16 of them. The distribution of attendance is bimodal. Most families either graduated or left within the first five sessions, 12.5 percent, and 9.2 percent attended between six and fifteen (Appendix Table B5). This bimodality is why the structural analysis treats participation as a binary graduation indicator rather than a continuous dose. The indicator separates families with nearly complete exposure from families with little exposure, and few families fall between the two. Appendix B.7 reports attendance and graduation cohort by cohort and describes the session and make-up rules in detail. These figures are computed on the experiment sample, and attendance is defined only for families offered the program, so, unlike the lottery assignment, it is not balanced by design.

Graduation is not random within the treatment group. Compared with non-graduates, graduates enter with higher baseline scores (8.3 percentage points on the JumpStart test, 5.1 on the Bracken) and a moderately more favorable home environment, in that they score higher on the economic well-being index, invest more at baseline, have markedly more children's books at home, and have fewer children in the household; child and parent demographics show no significant differences (Appendix Table B6). This selection on observables is precisely what the graduation equation of Section 5 absorbs, and Section 6.2 investigates the extent to which our structural

model can identify the importance of selection on unobservables.

# 4 Experimental Results

## 4.1 Evaluation Method

We use the following specification to estimate the effects of the JumpStart intervention on school readiness and parental investments:

$$Y_{i1} = \alpha + \beta_1 T_i + \beta_2 Y_{i0} + X_i\gamma + \tau_{s(i)} + \varepsilon_i, \tag{1}$$

where $Y_{i1}$ is child $i$'s outcome at endline, $T_i$ is the treatment group indicator for child $i$, which was assigned at the baseline, and $\varepsilon_i$ is an error term that is independent of $T_i$. We also control for the outcome at baseline $Y_{i0}$ and a set of demographic variables ($X_i$).[5] The term $\tau_{s(i)}$ is a fixed effect for the randomization stratum $s(i)$ to which child $i$ belongs. The main parameter we are interested in is $\beta_1$, which we interpret as the intention-to-treat (ITT) effect.

Alongside the conventional standard errors we report a randomization-inference $p$-value. We draw 2,000 alternative assignments, permuting $T_i$ within the randomization blocks so that each block keeps the number of treated families the lottery actually allocated to it, and re-estimate (1) on each draw. The reported $p$-value is the share of draws whose treatment coefficient, divided by its school-by-cohort clustered standard error, is at least as large in absolute value as the one obtained on the realized assignment.

## 4.2 The Impact of JumpStart on Test Scores and Parental Investment

Table 3 reports the intention-to-treat estimates for the three outcomes the remainder of the paper uses, across all four specifications; module-level results and the individual parental inputs are in Appendix Table C1, Panels A to C.

Control-group children learn substantially over the program year, gaining more than 18 percentage points on the JumpStart test, which is why the comparison group matters: any before-after comparison would mostly measure this growth. Relative to it, assignment to JumpStart raises the

[5] These demographics are measured at baseline and include the family's economic well-being index, parental years of schooling, an indicator of the parent being Hispanic, parent age, the number of children in the household, child age, and an indicator of a male child. The randomization strata are defined within cohort, so their fixed effects absorb the cohort indicators. Missing values are replaced by the sample mean of the variable, and we control for a single indicator equal to one when at least one of the controls is missing.

Table 3: Program Impacts at a Glance

| | Control mean | | Intention-to-Treat Effect | | | |
|---|---|---|---|---|---|---|
| | Baseline | Endline | (1) | (2) | (3) | (4) |
| JumpStart test score (% correct) | 50.45 | 68.70 | 7.41*** | 7.70*** | 7.58*** | 7.63*** |
| *Robust SE* | | | (1.38) | (1.24) | (1.13) | (1.11) |
| *School level SE* | | | (1.41) | (1.63) | (1.72) | (1.64) |
| *School-cohort level SE* | | | (1.32) | (1.48) | (1.39) | (1.33) |
| *Randomization inference p-value* | | | 0.00 | 0.00 | 0.00 | 0.00 |
| *N* | | | 890 | 890 | 890 | 890 |
| Bracken test score (% correct) | 23.05 | 33.96 | 3.11*** | 2.82** | 1.88* | 1.93* |
| *Robust SE* | | | (1.26) | (1.15) | (0.93) | (0.90) |
| *School level SE* | | | (1.11) | (1.30) | (1.11) | (1.13) |
| *School-cohort level SE* | | | (1.08) | (1.29) | (1.12) | (1.14) |
| *Randomization inference p-value* | | | 0.01 | 0.01 | 0.05 | 0.05 |
| *N* | | | 888 | 888 | 888 | 888 |
| Parental inputs factor score | 1.20 | 1.42 | 0.30*** | 0.27*** | 0.26*** | 0.26*** |
| *Robust SE* | | | (0.07) | (0.07) | (0.07) | (0.06) |
| *School level SE* | | | (0.07) | (0.08) | (0.07) | (0.07) |
| *School-cohort level SE* | | | (0.06) | (0.08) | (0.06) | (0.06) |
| *Randomization inference p-value* | | | 0.00 | 0.00 | 0.00 | 0.00 |
| *N* | | | 880 | 880 | 852 | 852 |
| Strata fixed effects | | | | ✓ | ✓ | ✓ |
| Baseline outcome | | | | | ✓ | ✓ |
| Demographic controls | | | | | | ✓ |

*Note:* Intention-to-treat estimates of assignment to JumpStart on the three outcomes the remainder of the paper uses: the overall JumpStart and Bracken scores, in percentage points, and the parental-investment factor score of Appendix B.4, in weekly-reading-hour units. Column (1) has no controls; (2) adds indicators for the randomization strata; (3) adds the baseline value of the outcome; (4) adds the baseline demographic controls. The first two columns report the control-group baseline and endline means, in percent correct for the tests and in weekly reading hours for the investment score. Each estimate is reported with robust, school-clustered, and school-by-cohort-clustered standard errors, and with the randomization-inference $p$-value from 2,000 permutations of the assignment within randomization blocks (Section 4). Stars refer to the school-by-cohort clustered $p$-value ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$). Module-level results and the remaining parental inputs are in Appendix Table C1, Panels A to C.

overall JumpStart score by 7.6 percentage points in the fully controlled specification, or 0.30 standard deviations of the baseline score, an estimate that is nearly identical across the four models and significant at the 1 percent level under every standard-error choice and under randomization inference. With 78.2 percent of offered families completing the program (Table 1), the Bloom ratio converts this offer effect into an effect of 9.8 percentage points on the families who complete when offered (Appendix Table E1). The gains are concentrated on the most curriculum-specific modules: 24.5 percentage points on name recognition and 21.5 on book handling, against 4.3 on colors and 4.8 on fine motor skills, with counting and sorting imprecisely estimated (Appendix

Table C1, Panel A). Although fewer than half of control children recognize their printed name at endline, about 70 percent of treated children do.

Effects on the Bracken test, which measures school readiness largely outside the program's curriculum, are much smaller and less robust. The estimate falls from 3.1 percentage points without controls to 1.9 in the fully controlled specification, or 0.10 standard deviations of the baseline score (randomization inference $p = 0.05$), and among the five modules only colors, the one whose content the curriculum covers directly, is significant at the 5 percent level (4.6 points; Appendix Table C1, Panel B). The program raises the national percentile score by 3.1 percentiles, significant at the 10 percent level, while the 3.0-point reduction in the share of children classified as delayed or very delayed is not significant. The module-level pattern gives only limited support to spillovers beyond the program's own curriculum, and Section 4.5 shows the overall Bracken result is also not robust to attrition.

The program also changes what parents do at home. The parental-investment factor score, the index that aggregates the survey's investment items and serves as the investment measure $X_{i1}$ in the model of Section 5, rises by 0.26 points, or 0.26 standard deviations of its endline distribution in the control group (randomization inference $p < 0.001$), an effect stable across specifications and larger than the 0.22-point growth the control group experiences on its own. The item-level estimates show where investment moves. Treated families read to their child on 0.56 more days per week, although the weekly hours of reading the index uses rise by only 0.22, significant at the 10 percent level, and the number of children's books at home rises by 1.27, an effect likely attributable to the dozen books the curriculum supplies; the remaining inputs, including television exposure and grocery trips, show no effects, unsurprising for the help-with-skills activities that are nearly universal in both groups at endline (Appendix Table C1, Panel C).

Can this rise in investment, by itself, account for the test-score gains? Section 5 builds and estimates a model in which graduation can raise knowledge directly and through the investment it induces, with the factor score serving as the investment measure $X_{i1}$, and delivers the decomposition of the program's effect into the two channels.

### 4.3 Parental Beliefs

A natural candidate channel for a parent-directed program is a change in what parents believe about the productivity of their investments. We examine one measure of this channel in the third cohort, eliciting beliefs at baseline and endline through five pairs of vignette questions in which two hypothetical three-year-olds receive 30 versus 10 minutes per day of the program's core ac-

tivities, and scoring whether the respondent expects the more-invested child to reach a higher developmental level; Appendix C.4 details the instrument and the two scores built from it. Beliefs show moderate persistence, with a coefficient on the baseline score of about one third, and we detect no effect of assignment to JumpStart on these elicited scores in any specification (Appendix Table C2). For parsimony, the model of Section 5 omits a beliefs channel.

### 4.4 Heterogeneous Treatment Effects

We test whether the program's effects vary with baseline characteristics by splitting the experiment sample at the median of six baseline variables (the two test scores, child age, the economic well-being index, parental age, and parental schooling), and re-estimating the fully controlled specification within each subgroup for the JumpStart score, the Bracken score, and the investment factor score; Appendix C.5 reports every estimate with equality tests. The evidence for heterogeneity is limited. On the JumpStart test, every subgroup estimate is positive and significant, ranging from 5.8 to 11.4 percentage points around the full-sample 7.6, and no equality test rejects at the 10 percent level. On the investment factor score, the effects are positive in every subgroup, and no split detects statistically significant heterogeneity. The largest contrast on the program's own curriculum is by parental education, with gains of 11.4 percentage points among below-median-schooling parents against 6.2 among the rest ($p = 0.19$), and the splits by baseline test scores point in the same direction. This gradient is imprecise in the subgroup estimates, but it reappears inside the structural model, where the return to graduation is allowed to vary with parental schooling and baseline knowledge and both interactions are negative and significant, so the non-structural and structural halves of the paper agree on where the program works.

### 4.5 Attrition

Endline response was higher in the treatment group, particularly in the first cohort, and attriters differ from stayers on several observables, in that they are younger and less likely to be Hispanic, and their children score slightly higher on the baseline Bracken test (Appendix C.6). We therefore re-estimate the ITT effects under four corrections: inverse probability weighting, a Heckman selection model, Lee bounds, and a balancing procedure that matches treatment and control families on the estimated attrition propensity within randomization strata and drops any pair with an attriting member; Appendix C.6 details the procedure and reports all outcomes. The JumpStart effect is robust, ranging from 7.1 to 7.6 percentage points across all point-estimate corrections and remaining significant at the 1 percent level even at the Lee lower bound of 3.6 percentage points.

The investment factor score is less robust. It survives the three point-estimate corrections, at 0.19 to 0.26 and significant at the 5 percent level, but not the Lee lower bound, with Lee bounds of [0.07, 0.38], and the effect on reading frequency behaves the same way. The effect on children's books survives every correction, including both Lee bounds. The Bracken effect is not robust, since it loses significance under the balancing procedure, and its Lee lower bound is -1.0 percentage points.

# 5 The Mechanisms

The experimental results of Section 4 establish two facts. First, children whose parents were offered the JumpStart program score higher at the end of the year, and second, parents who were offered the program invest more at home. They do not establish how the two are connected. The gain in knowledge may come from the program's own contact with the household. The liaison works with the children directly once a month and teaches parents how to deliver the specific content of the curriculum, instruction that can raise scores without changing how much parents invest. Alternatively, the gain could be generated by the additional investment that graduation induces, which raises knowledge through the production technology. The distinction matters for policy, because the two channels have different horizons. The program's own contact ends with the sessions, whereas a change in what parents do at home can persist once the routine is established, and the program's effect persists with it.

The experiment alone cannot separate the channels, because assignment moves both at once. Separation requires assumptions beyond randomization, and this section makes them explicit within a model of skill formation, parental investment, and program participation, which it specifies, identifies from the experiment, and estimates. The model delivers the average effect of graduation, its decomposition into a direct and an investment channel, and the effects on the treated and the untreated. Section 6 then establishes which of these conclusions depend on the model's structure and which do not.

## 5.1 The Model

### 5.1.1 The Environment

The model describes a single school year in the life of a household made up of one parent and one child. Each household $i$ is randomly assigned to the control group ($T_i = 0$) or to the treatment group ($T_i = 1$), and assignment confers access to the JumpStart program, which is

the only route into it. Parents in the treatment group decide whether to complete it, which we call *graduating* ($G_i = 1$); the term refers to the parents' completion of the program. Parents in the control group cannot graduate. Letting $G_i^*$ denote the net payoff to graduation, the realized participation indicator is

$$G_i = T_i \cdot \mathbf{1}\{G_i^* \geq 0\}, \tag{2}$$

so that the experiment has one-sided compliance, with $T_i = 0$ implying $G_i = 0$ with certainty. The net payoff $G_i^*$ is derived from the parent's problem in Section 5.2.2. Throughout, a parent in the treatment group, to whom the program has been offered, has $T_i = 1$, a graduate has $G_i = 1$, and a never-taker has $G_i^* < 0$. A control parent with observed $G_i = 0$ need not be a never-taker.

The child enters the year with knowledge $K_{i0}$ and leaves it with knowledge $K_{i1}$, while the parent enters with an investment routine $X_{i0}$ and chooses investment $X_{i1}$ during the year. Events unfold in a fixed order. Assignment $T_i$ is realized first, after which parents in the treatment group observe their taste shocks and decide whether to graduate. Investment $X_{i1}$ follows, and end-of-year knowledge $K_{i1}$ is produced last.

### 5.1.2 The Technology of Knowledge Production

End-of-year knowledge is produced according to

$$K_{i1} = \beta_0 + \beta_{K_0} K_{i0} + \beta_{X_1} X_{i1} + \big(\beta_G + \boldsymbol{\beta}_W' \boldsymbol{W}_i + \varepsilon_{i1}\big) G_i + \boldsymbol{\beta}_{Z_K}' \boldsymbol{Z}_{K,i} + \varepsilon_i. \tag{3}$$

Equation (3) follows a value-added specification (see e.g., Todd and Wolpin, 2007). Knowledge cumulates, in that the stock the child holds at the start of the year carries forward at rate $\beta_{K_0}$, and the remaining terms are the knowledge added over the year, produced by contemporaneous investment $X_{i1}$, by graduation $G_i$, and by a vector $\boldsymbol{Z}_{K,i}$ of characteristics that enter the technology directly.

Three features of the technology matter for what follows. First, $\beta_{X_1}$, the marginal product of investment, represents the rate at which home investment is converted into knowledge.[6] Second, $\beta_G$ is the effect of graduation holding measured investment fixed. It captures what graduation adds to knowledge without passing through measured investment, such as the direct contact between the child and the liaison at the monthly sessions children attend, the program materials other than books, and any other input that graduation raises but that our investment measure does not

[6] We do not allow $\beta_{X_1}$ to differ by graduation status, that is, we exclude a $G_i X_{i1}$ interaction. Our descriptive analysis shows no detectable pattern of this kind, and an earlier version of the model that included the interaction found no evidence for it.

record. We refer to it as the *direct* effect, understanding throughout that it also absorbs those unmeasured inputs. Third, the return to graduation is allowed to vary with observed characteristics $\boldsymbol{W}_i$, through $\boldsymbol{\beta}_W$, and with an unobserved idiosyncratic gain $\varepsilon_{i1}$ (normalized to $\mathbb{E}[\varepsilon_{i1}] = 0$). The presence of $\varepsilon_{i1}$ is what makes the graduation decision a selection problem, because parents who expect to gain more may be more likely to graduate.

### 5.1.3 Preferences

Parents' utility depends on their child's knowledge and their own investment. We represent preferences by the utility function

$$U_i(X_{i1}, G_i) = \kappa K_{i1} + \gamma_{1i} X_{i1} - \tfrac{1}{2}\big(X_{i1} - \gamma_{2i}\big)^2 + \big(\alpha_0 + \boldsymbol{\alpha}'_{Z_G} \boldsymbol{Z}_{G,i} - \eta_i\big) G_i, \tag{4}$$

where $\kappa$ is the marginal utility of the child's knowledge and $\eta_i$ an idiosyncratic distaste for completing the program. The quadratic term penalizes departures from a parent-specific investment target $\gamma_{2i}$, and $\gamma_{1i}$ is the parent's marginal taste for investment. Both are allowed to depend on observable characteristics, and the marginal taste additionally carries an unobserved shock,

$$\gamma_{1i} = \gamma_{10} + \gamma_G G_i + \boldsymbol{\gamma}'_{Z_X} \boldsymbol{Z}_{X,i} + \nu_i, \qquad \gamma_{2i} = \gamma_{20} + \gamma_{X_0} X_{i0}, \tag{5}$$

where $\boldsymbol{Z}_{X,i}$ is a vector of baseline characteristics and $\nu_i$, normalized to $\mathbb{E}[\nu_i] = 0$, is an unobserved shock to the parent's marginal taste for investment. It stands for the part of a family's willingness to invest that the observed characteristics do not record. It is also what makes investment endogenous in the graduation decision, since the same shock enters the payoff to graduating through the term $\gamma_G \nu_i$ of (11), so a parent with an unusually high taste for investment both invests more and, when $\gamma_G > 0$, gains more from graduating.

The parameter $\gamma_{X_0}$ lets the family's investment target depend on its baseline investment. A positive $\gamma_{X_0}$ means that families who invested more at baseline invest more at endline, holding covariates fixed, which generates *inertia* in parental investment. Inertia of this kind is consistent both with routines that persist in their own right and with persistent family characteristics the covariates do not absorb, and we do not attempt to separate the two.[7] The parameter $\gamma_G$ is the shift in the marginal taste for investment caused by graduation. It is the sole channel through which the program alters how much parents invest.

[7] Specifications in which current choices are anchored to past behavior go back to Duesenberry (1949); see also Ryder and Heal (1973), Becker and Murphy (1988), and Constantinides (1990).

The specification contains no budget constraint. Under quasi-linearity, a linear cost of investment time or materials enters the first-order condition exactly as the marginal taste for investment $\gamma_{1i}$ does. As a result, such costs are absorbed into the taste parameters and not separately identified. In our specification, we account for variation in family resources through the economic well-being index included in $\boldsymbol{Z}_{X,i}$. The costs that bear on the graduation decision in this setting are the time and travel costs of attending the sessions, and we account for them through the observed characteristics in $\boldsymbol{Z}_{G,i}$, which enter the graduation decision.

## 5.2 Model Solution

The model is solved by backward induction. The parent first chooses the investment that maximizes utility under each graduation status, taking that status as given. Substituting the resulting investment back into the utility function delivers the payoff to graduating and to not graduating, and the comparison of the two yields the graduation decision rule. The two steps are taken in that order below.

### 5.2.1 Optimal Investment

Parents choose $X_{i1} \geq 0$ to maximize their utility (4). With strict concavity in $X_{i1}$, the first-order condition yields the optimal investment choice:[8]

$$X_{i1}^{*} = \kappa\beta_{X_1} + \gamma_{1i} + \gamma_{2i}. \tag{6}$$

The parent invests up to the point where the marginal cost of departing from her target ($\gamma_{2i}$) equals the marginal taste for investment ($\gamma_{1i}$) plus the marginal value of the knowledge it produces, $\kappa\beta_{X_1}$. Substituting (5) into (6) and collecting terms gives the optimal investment equation

$$X_{i1}^{*} = \widetilde{\gamma}_0 + \gamma_{X_0} X_{i0} + \gamma_G G_i + \boldsymbol{\gamma}_{Z_X}' \boldsymbol{Z}_{X,i} + \nu_i, \tag{7}$$

[8]The parents' utility maximization problem admits a corner solution for certain values of the parameters. Appendix D.1 discusses this possibility. We assume an interior solution for both graduation statuses. The investment factor score cannot reveal a corner, since its origin is set by the normalization of Appendix B.4, so the sign of a score carries no information about one. The items from which the score is built nonetheless suggest that corners are rare, since at endline a single family among the 887 with all three items recorded reports no reading, no children's books, and no help with school-readiness skills.

where the composite constant is defined as[9]

$$\widetilde{\gamma}_0 \equiv \gamma_{10} + \gamma_{20} + \kappa\beta_{X_1}. \tag{8}$$

While equation (7) generally applies to every household, for parents in the control group the graduation term drops out since $G_i = 0$ by one-sided compliance (2), so their optimal investment is simply

$$X_{i1}^* = \widetilde{\gamma}_0 + \gamma_{X_0} X_{i0} + \boldsymbol{\gamma}'_{Z_X} \boldsymbol{Z}_{X,i} + \nu_i. \tag{9}$$

Graduation therefore shifts investment by exactly $\gamma_G$, uniformly across households.

### 5.2.2 The Graduation Decision

A treatment-arm parent graduates when the utility of doing so, evaluated at the investment she would optimally choose in each case, is at least as large as the utility of not graduating. Substituting (7) into (4) and collecting the terms that multiply $G_i$ we can write the net payoff to graduation as a linear function in observables and a single composite error:

$$G_i^* = \widetilde{\alpha}_0 + \widetilde{\alpha}_{X_0} X_{i0} + \widetilde{\boldsymbol{\alpha}}'_{Z_X} \boldsymbol{Z}_{X,i} + \boldsymbol{\alpha}'_{Z_G} \boldsymbol{Z}_{G,i} + \widetilde{\boldsymbol{\alpha}}'_W \boldsymbol{W}_i + \omega_i, \tag{10}$$

with

$$\omega_i \equiv \kappa\varepsilon_{i1} + \gamma_G \nu_i - \eta_i, \qquad \widetilde{\alpha}_{X_0} = \gamma_{X_0}\gamma_G, \qquad \widetilde{\boldsymbol{\alpha}}_{Z_X} = \gamma_G \boldsymbol{\gamma}_{Z_X}, \qquad \widetilde{\boldsymbol{\alpha}}_W = \kappa\boldsymbol{\beta}_W, \tag{11}$$

while the composite intercept $\widetilde{\alpha}_0$ is given in Appendix D. For a household in the treatment group, the rule (2) therefore reduces to

$$G_i = \mathbf{1}\{G_i^* \geq 0\},$$

with $G_i^*$ given by equation (10).

Each composite term in (11) has a direct interpretation. First, baseline investment and the taste covariates enter the graduation index only through $\gamma_G$, as parental prior routine matters for the graduation choice only because the latter will change investments at endline. If graduation left the taste for investment unchanged, so that $\gamma_G = 0$, then $\widetilde{\alpha}_{X_0} = \widetilde{\boldsymbol{\alpha}}_{Z_X} = \mathbf{0}$ and baseline investment would drop out of the decision rule entirely, and the taste covariates would enter it only through the participation costs in $\boldsymbol{Z}_{G,i}$. Similarly, $\boldsymbol{W}_i$ enters the index only through $\kappa\boldsymbol{\beta}_W$, as observed characteristics shift the graduation decision by shifting the anticipated return, valued at the marginal

[9] Throughout, a tilde marks a coefficient that is a composite of the structural parameters rather than a structural parameter itself.

utility of knowledge. The selection error $\omega_i$ inherits the unobserved gain $\varepsilon_{i1}$, so, for $\kappa > 0$, parents who anticipate larger gains are more likely to graduate.

### 5.2.3 Knowledge Under Optimal Investment

Substituting optimal investment (7) into the technology (3) we obtain a reduced-form knowledge equation as a function of baseline knowledge, baseline investment, graduation, and the covariates:

$$K_{i1} = \widetilde{\beta}_0 + \beta_{K_0} K_{i0} + \widetilde{\beta}_{X_0} X_{i0} + \big(\widetilde{\beta}_G + \boldsymbol{\beta}'_W \boldsymbol{W}_i + \varepsilon_{i1}\big) G_i + \widetilde{\boldsymbol{\beta}}'_{X,K} \boldsymbol{Z}_{X,K,i} + \xi_i, \tag{12}$$

where

$$\widetilde{\beta}_0 = \beta_0 + \beta_{X_1} \widetilde{\gamma}_0, \qquad \widetilde{\beta}_{X_0} = \beta_{X_1} \gamma_{X_0}, \qquad \widetilde{\beta}_G = \beta_G + \beta_{X_1} \gamma_G, \qquad \xi_i \equiv \beta_{X_1} \nu_i + \varepsilon_i, \tag{13}$$

and $\boldsymbol{Z}_{X,K,i}$ collects each distinct covariate appearing in $\boldsymbol{Z}_{X,i}$ or $\boldsymbol{Z}_{K,i}$, which implies $\widetilde{\boldsymbol{\beta}}'_{X,K} \boldsymbol{Z}_{X,K,i} \equiv \beta_{X_1} \boldsymbol{\gamma}'_{Z_X} \boldsymbol{Z}_{X,i} + \boldsymbol{\beta}'_{Z_K} \boldsymbol{Z}_{K,i}$.

The composite terms in (13) show what a reduced-form equation for end-of-year knowledge can recover. The graduation effect $\widetilde{\beta}_G$ bundles two channels, a direct effect $\beta_G$ and an indirect effect $\beta_{X_1}\gamma_G$, the knowledge produced by the investment that graduation induces. Section 5.4 shows that separating the two relies on baseline investment $X_{i0}$ affecting endline knowledge only through the investment $X_{i1}$ it induces, so that it acts as an exclusion restriction for the investment equation.

## 5.3 Measurement

We measure $K_{i0}$ and $K_{i1}$ by the JumpStart test score at baseline and endline, expressed as the percentage of items answered correctly, and $G_i$ by the graduation indicator. We measure $X_{i0}$ and $X_{i1}$ by factor scores extracted from the parental-investment items collected in each survey wave, using regression scores for $X_{i0}$ and Bartlett scores for $X_{i1}$, following Skrondal and Laake (2001).[10] The investment index $X_{it}$ is the factor score built from the five parental-investment items described in Section 3.2 and detailed in Appendix B.3. Appendix B.4 describes the factor model and the scaling of the resulting scores.

[10] Write the Bartlett score of endline investment as $\widehat{X}^B_{i1} = X_{i1} + \epsilon^B_{i1}$ and the regression score of baseline investment as $\widehat{X}^R_{i0} = X_{i0} - \epsilon^R_{i0}$, where $\epsilon^B_{i1}$ is conditionally mean zero and $\epsilon^R_{i0}$ is uncorrelated with $\widehat{X}^R_{i0}$, so that the first error only adds to the disturbance and the second causes no attenuation. See Appendix B.4 for a full discussion.

The vector $\boldsymbol{W}_i$ that governs observed heterogeneity in the return to graduation contains an indicator for above-median baseline knowledge and an indicator for above-median parental schooling. The control vectors $\boldsymbol{Z}_{X,i}$ and $\boldsymbol{Z}_{K,i}$ are identical. Each contains the index of the family's economic well-being, the number of children in the household, parental age, parental years of schooling, an indicator for Hispanic parents, cohort indicators, the child's age at the test, and an indicator for a male child.

The participation-cost vector $\boldsymbol{Z}_{G,i}$ contains the same controls and adds the driving distance, in kilometers, between the family's home and the campus at which the program was held, constructed as described in Appendix B.6. Distance raises the cost of attending and completing the program and therefore shifts the graduation decision. We exclude it from $\boldsymbol{Z}_{X,i}$ and $\boldsymbol{Z}_{K,i}$, on the grounds that, conditional on the controls, how far a family lives from the campus bears neither on the marginal taste for reading to one's child at home nor on the technology by which investment is converted into knowledge.

### 5.4 Identification

The model is identified by the parametric restrictions imposed above together with two additional assumptions. Let $\boldsymbol{Z}_i$ collect a constant and all distinct observed covariates appearing in $X_{i0}$, $\boldsymbol{Z}_{X,i}$, $\boldsymbol{Z}_{K,i}$, $\boldsymbol{Z}_{G,i}$, and $\boldsymbol{W}_i$. The unobserved components of the model are then the shock to the technology $\varepsilon_i$, the idiosyncratic gain from graduating $\varepsilon_{i1}$, the shock to the marginal taste for investment $\nu_i$, and the distaste for completing the program $\eta_i$.

**Assumption 1 (Random assignment).** *Conditional on $\boldsymbol{Z}_i$, the offer is independent of every unobserved component of the model,*

$$\left(\varepsilon_i, \varepsilon_{i1}, \nu_i, \eta_i\right) \;\perp\; T_i \;\middle|\; \boldsymbol{Z}_i.$$

**Assumption 2 (Joint normality).** *Conditional on $\boldsymbol{Z}_i$, the unobserved components are jointly normal with mean zero and with a covariance matrix $\boldsymbol{\Sigma}$,*

$$\left(\varepsilon_i, \varepsilon_{i1}, \nu_i, \eta_i\right) \,\middle|\, \boldsymbol{Z}_i \;\sim\; \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right).$$

The lottery supports Assumption 1 within school and randomization block (Section 3.1). Separately, the absence of $T_i$ from the investment rule and knowledge technology maintains that an offer affects these outcomes only through graduation.

Two consequences carry the identification argument. The first concerns what an offer does when it does not lead to graduation. Since $T_i$ appears in neither the technology (3) nor the investment

rule (7), the offer affects knowledge and investment only through $G_i$, and Assumption 1 adds that it is independent of their shocks. Conditional on $\boldsymbol{Z}_i$, an offer that does not lead to graduation then leaves endline knowledge, the gain from graduating, the marginal taste for investment, and the net payoff to graduation $G_i^*$ exactly where they would have been without it.[11] The second consequence concerns the control arm. Control families cannot graduate, and randomization makes them representative of the whole population, so they identify the investment rule and the technology at $G_i = 0$ with no correction for selection. Section 6.1 takes up the exclusion and the randomization separately, since the nonparametric bounds rest on the same two restrictions, and gives the evidence for the exclusion.

Assumption 2 restricts the distribution of the shocks given the observables, and together with Assumption 1 it fixes that distribution within each arm, which is what the control functions below are built on. Section 6 assesses, for each conclusion, how much depends on that parametric structure, and provides partial-identification inference for the quantities that do. Identification proceeds one equation at a time, and Appendix D gives the derivations.

**Graduation equation.** Under Assumption 2, $\omega_i$ is normal with standard deviation $\sigma_\omega$, and under Assumption 1 it is independent of $T_i$ conditional on $\boldsymbol{Z}_i$, so the graduation rule (2)–(10) implies that graduation in the treatment arm follows a probit,

$$\Pr(G_i = 1 \mid T_i = 1, \boldsymbol{Z}_i) = \Phi(\boldsymbol{a}'\boldsymbol{Z}_i), \tag{14}$$

with $\Phi(\cdot)$ the cdf of a standard normal distribution and

$$a_0 = \frac{\widetilde{\alpha}_0}{\sigma_\omega}, \qquad a_{X_0} = \frac{\widetilde{\alpha}_{X_0}}{\sigma_\omega}, \qquad a_j = \frac{\widetilde{\alpha}_{Z_X,j} + \alpha_{Z_G,j} + \widetilde{\alpha}_{W,j}}{\sigma_\omega}.$$

The last expression holds for every other covariate $j$ in $\boldsymbol{Z}_i$, with a coefficient set to zero when $j$ does not enter the corresponding vector. The probit identifies each $a_j$, which is a composite of structural parameters whenever $j$ enters more than one term of the graduation payoff and is zero when $j$ enters only $\boldsymbol{Z}_{K,i}$. Control households carry no information on this equation, since they cannot graduate.

**Investment equation.** Graduating parents are selected on $\omega_i$, which carries their taste shock $\nu_i$ through the term $\gamma_G \nu_i$ of (11), so the mean of $\nu_i$ among graduates is not zero. Conditioning (7) on

[11] The statement concerns an offer that does not end in graduation. The offer as a whole does raise knowledge, which is the intention-to-treat effect, but only by giving access to the program, and hence through graduation.

$(\boldsymbol{Z}_i, G_i)$ therefore leaves a term in

$$\lambda_i \;\equiv\; \mathbb{E}\Big[\tfrac{\omega_i}{\sigma_\omega} \;\Big|\; T_i = 1, G_i, \boldsymbol{Z}_i\Big], \tag{15}$$

the control function, which is the average selection error among families with the same observables and the same graduation status, in units of its own standard deviation. It is a non-linear function of the graduation index, with the closed form under Assumption 2 given in (20) below. The estimating equation is

$$X_{i1} = \widetilde{\gamma}_0 + \gamma_{X_0} X_{i0} + \gamma_G G_i + \boldsymbol{\gamma}'_{Z_X} \boldsymbol{Z}_{X,i} + \rho_\nu T_i \lambda_i + e_{X,i}, \tag{16}$$

with $\rho_\nu = \mathrm{Cov}(\nu_i, \omega_i)/\sigma_\omega$ measuring how strongly the taste for investment and the taste for graduating move together. It is written for the pooled sample, and the correction carries $T_i$ because a control family faces no graduation decision, so its taste shock is untruncated and $\mathbb{E}[\nu_i \mid T_i = 0, \boldsymbol{Z}_i] = 0$. Least squares on (16) identifies every coefficient of (7), namely the composite intercept $\widetilde{\gamma}_0$, the inertia parameter $\gamma_{X_0}$, the graduation shift $\gamma_G$, the taste loadings $\boldsymbol{\gamma}_{Z_X}$, and the selection coefficient $\rho_\nu$ on the correction itself. The control arm identifies $(\widetilde{\gamma}_0, \gamma_{X_0}, \boldsymbol{\gamma}_{Z_X})$ on its own, so pooling the two arms over-identifies them.

**Knowledge equation.** Conditioning the reduced form (12) in the same way requires two corrections rather than one. The idiosyncratic gain $\varepsilon_{i1}$ enters the knowledge of graduates only, so the error of the knowledge equation covaries with the selection error differently for graduates and non-graduates, and each group takes its own coefficient on the control function. Writing (15) as $\lambda_{1i}$ for graduates and $\lambda_{0i}$ for non-graduates, with closed forms given in (20), the estimating equation is

$$\begin{aligned} K_{i1} = \widetilde{\beta}_0 + \beta_{K_0} K_{i0} + \widetilde{\beta}_{X_0} X_{i0} + \big(\widetilde{\beta}_G + \boldsymbol{\beta}'_W \boldsymbol{W}_i\big) T_i G_i + \widetilde{\boldsymbol{\beta}}'_{X,K} \boldsymbol{Z}_{X,K,i} \\ + \rho_0 T_i (1 - G_i) \lambda_{0i} + \rho_1 T_i G_i \lambda_{1i} + e_{K,i}. \end{aligned} \tag{17}$$

Least squares on (17) identifies $\big(\widetilde{\beta}_0,\ \beta_{K_0},\ \widetilde{\beta}_{X_0},\ \widetilde{\beta}_G,\ \boldsymbol{\beta}_W,\ \widetilde{\boldsymbol{\beta}}_{X,K}\big)$, together with the selection coefficients $\rho_0$ for non-graduates and $\rho_1$ for graduates.

The structural parameters are obtained by inverting the maps (11) and (13). The marginal product of investment is

$$\beta_{X_1} = \frac{\widetilde{\beta}_{X_0}}{\gamma_{X_0}}. \tag{18}$$

Baseline investment therefore serves as an exclusion restriction for endline investment, and the decomposition rests on the two properties any exclusion restriction must have.[12]

Relevance requires $\gamma_{X_0} \neq 0$. Investment behavior must persist, because without inertia baseline investment predicts nothing about endline investment, the denominator of (18) vanishes, and the marginal product of investment cannot be recovered. Validity requires that $X_{i0}$ be absent from the technology (3), which is the value-added restriction. Were baseline investment to raise endline knowledge directly, that direct effect would load onto $\widetilde{\beta}_{X_0}$, and $\beta_{X_1}$ would confound the marginal product of investment with it. The first property is an estimated feature of the data, reported below. The second is an assumption the data cannot test, and Section 6.3 quantifies how far the decomposition moves as it is relaxed.

Once $\beta_{X_1}$ is identified, we can recover the direct effect of program graduation:

$$\beta_G = \widetilde{\beta}_G - \beta_{X_1}\gamma_G, \tag{19}$$

while the intercept is given by $\beta_0 = \widetilde{\beta}_0 - \beta_{X_1}\widetilde{\gamma}_0$. Equations (18) and (19), together with $\gamma_G$ from the investment equation and $\boldsymbol{\beta}_W$ from the knowledge equation, are all that the treatment effects require. The scale of the selection error follows from $\sigma_\omega = \gamma_{X_0}\gamma_G/a_{X_0}$. The probit delivers $a_{X_0}$, the coefficient on baseline investment in the normalized index, while the investment equation delivers the same coefficient in the unnormalized index, $\widetilde{\alpha}_{X_0} = \gamma_{X_0}\gamma_G$, because the utility function (4) is written in units of $X_{i1}$. Their ratio is the normalizing constant, which de-normalizes $G_i^*$ and puts the graduation payoff in the same units as investment. The remaining utility parameters are then identified whenever $\gamma_G \neq 0$, with two exceptions. The marginal utility of knowledge $\kappa$ is identified only when the return to graduation varies with $\boldsymbol{W}_i$ (Appendix D.5), and so is the utility intercept $\alpha_0$, which depends on it. Appendix D gives these steps and notes which parameters are identified only as composites.

**Control functions.** Assumptions 1 and 2 make $\omega_i/\sigma_\omega$ standard normal conditional on $\boldsymbol{Z}_i$ and $T_i$, so that $G_i = \mathbf{1}\{\boldsymbol{a}'\boldsymbol{Z}_i + \omega_i/\sigma_\omega \geq 0\}$ and the two branches of (15) are the inverse Mills ratios

$$\lambda_{1i} = \frac{\phi(\boldsymbol{a}'\boldsymbol{Z}_i)}{\Phi(\boldsymbol{a}'\boldsymbol{Z}_i)}, \qquad \lambda_{0i} = \frac{-\phi(\boldsymbol{a}'\boldsymbol{Z}_i)}{1-\Phi(\boldsymbol{a}'\boldsymbol{Z}_i)}, \qquad \lambda_i = G_i\lambda_{1i} + (1-G_i)\lambda_{0i}, \tag{20}$$

[12]Equation (18) is indirect least squares, being the ratio of two reduced-form coefficients, and it has the same probability limit as the two-stage least squares coefficient from a regression of endline knowledge on endline investment, graduation and its interactions with $\boldsymbol{W}_i$, the controls, and the two control-function terms of (17), using baseline investment as the instrument. The system is just identified, since every other shifter in (7) also enters the technology (3).

with $\phi(\cdot)$ the standard normal density and $\boldsymbol{a}'\boldsymbol{Z}_i$ the index of the probit (14). Appendix D.4 derives them.

The normal form of these control functions is what identifies the selection coefficients. Equations (16) and (17) are linear regressions, and $\lambda_i$ can be told apart from the controls only because it is a curved function of the index in which the controls enter linearly. This requires no exclusion restriction, but we impose one. Home-to-campus driving distance enters the graduation payoff (10) but neither the technology (3) nor the investment rule (7), so it shifts a family's chance of graduating without shifting its knowledge or its investment directly. Such a shifter widens the spread of the estimated graduation probabilities, and the wider that spread, the less $\lambda_i$ resembles a linear function of the controls.[13]

### 5.5 Treatment Effects and Their Decomposition

Differencing the technology (3) at $G_i = 1$ and $G_i = 0$, and replacing investment by the optimal rule (7), gives the individual treatment effect of graduation on child $i$'s end-of-year knowledge:

$$\Delta_i = \underbrace{\beta_G + \boldsymbol{\beta}_W'\boldsymbol{W}_i}_{\text{direct channel}} + \underbrace{\beta_{X_1}\gamma_G}_{\substack{\text{investment}\\ \text{channel}}} + \underbrace{\varepsilon_{i1}}_{\substack{\text{unobserved}\\ \text{gain}}} . \tag{21}$$

Equation (21) splits the effect of graduating into three parts. The direct channel collects what the JumpStart sessions contribute at unchanged measured investment, together with whatever graduation changes among inputs the investment index does not record. The investment channel is a product of two estimated quantities: graduation raises the taste for investment by $\gamma_G$, and each unit of investment produces $\beta_{X_1}$ points of knowledge. The two channels can therefore be switched off one at a time, since $\gamma_G = 0$ removes the investment channel and $\beta_G = 0$ together with $\boldsymbol{\beta}_W = \mathbf{0}$ removes the direct one. Because $\beta_{X_1}$ does not vary with graduation status, the investment channel is common to all households, and observed heterogeneity operates entirely through the direct channel.

Averaging (21) over the offered families, over those who graduate, and over offered non-graduates (never-takers) gives the average treatment effect (ATE), the effect on the treated (ATT), and the

[13] In our data distance shifts graduation only weakly, so the spread it adds is narrow and the selection parameters are weakly identified (Section 5.7 and Appendix F.1).

effect on the untreated (ATU),

$$\mathrm{ATE} = \beta_G + \boldsymbol{\beta}_W' \mathbb{E}[\boldsymbol{W}_i] + \beta_{X_1}\gamma_G, \tag{22}$$

$$\mathrm{ATT} = \beta_G + \boldsymbol{\beta}_W' \mathbb{E}[\boldsymbol{W}_i \mid T_i = 1, G_i = 1] + \beta_{X_1}\gamma_G + (\rho_1 - \rho_0)\,\mathbb{E}[\lambda_{1i} \mid T_i = 1, G_i = 1], \tag{23}$$

$$\mathrm{ATU} = \beta_G + \boldsymbol{\beta}_W' \mathbb{E}[\boldsymbol{W}_i \mid T_i = 1, G_i = 0] + \beta_{X_1}\gamma_G + (\rho_1 - \rho_0)\,\mathbb{E}[\lambda_{0i} \mid T_i = 1, G_i = 0]. \tag{24}$$

The effects on the treated and the untreated differ from the average effect for two reasons. The composition of $\boldsymbol{W}_i$ differs across the two groups, and unobserved gains may be correlated with the decision to graduate. The latter is governed by the *selection wedge*

$$\rho_1 - \rho_0 = \frac{\mathrm{Cov}(\varepsilon_{i1}, \omega_i)}{\sigma_\omega}, \tag{25}$$

which is zero precisely when parents do not sort into graduation on the basis of their idiosyncratic gain. Finally, one-sided compliance implies the well-known result of Bloom (1984)[14]

$$\begin{aligned} \mathrm{ITT} &= \Pr(G_i = 1 \mid T_i = 1) \cdot \mathrm{ATT} \\ &\equiv \pi \cdot \mathrm{ATT} \end{aligned} \tag{26}$$

where $\pi \equiv \mathbb{E}[p_i \mid T_i = 1]$ is the graduation rate among offered families, the average over them of the propensity score $p_i \equiv \Phi(\boldsymbol{a}'\boldsymbol{Z}_i)$, the probability (14) that family $i$ graduates when offered. The model-implied intention-to-treat effect can therefore be compared directly with the experimental estimate of Section 4.

### 5.6 Estimation

Estimation mirrors identification. We first estimate the probit (14) on the treatment-group sample and evaluate the control functions (20) at the fitted index. We then estimate the investment and knowledge equations (16) and (17) jointly on the pooled sample by seemingly unrelated regression, allowing their errors to be correlated within a household. Inverting the reduced-form maps of Section 5.4 recovers the structural parameters. Evaluating (22) to (24) at those estimates, together with the sample means of $\boldsymbol{W}_i$, $\lambda_{0i}$, and $\lambda_{1i}$ over the relevant group, delivers the treatment effects of Section 5.5. We report two sets of standard errors, from the delta method with the

[14] The derivation of this equality is shown in Appendix E.3. One-sided compliance also makes the effect on the treated coincide with the local average treatment effect of Imbens and Angrist (1994). With no always-takers, the compliers are exactly the families who graduate when offered, so the population the instrument identifies and the population of graduates are the same.

underlying reduced-form standard errors clustered by cohort and school, and from a bootstrap of the full estimation procedure that resamples the school-by-cohort clusters (2,000 replications) and so also accounts for the estimated control functions. The two are very similar, and we note where they differ. The model is estimated on a sample of $N$ = 842 families, of whom 454 are in the treatment group and 388 in the control group.[15]

### 5.7 Estimation Results

Table 4 reports the structural estimates and Table 5 the implied treatment effects. Column (1) restricts the return to graduation to be homogeneous in observables, imposing $\boldsymbol{\beta}_W = \mathbf{0}$; column (2) allows it to vary with $\boldsymbol{W}_i$. Each specification occupies an estimate column and two standard-error columns, delta method and bootstrap, in both tables, and we refer to it by its number. Both specifications use home-to-campus driving distance as the exclusion restriction for the graduation choice and condition on the controls listed in Section 5.3, with missing values replaced by the sample mean and flagged, as in the intention-to-treat regressions, by a single indicator equal to one when at least one control is missing. Column (2) additionally enters the two components of $\boldsymbol{W}_i$ as controls in the production function, so that $\boldsymbol{\beta}_W$ measures how the return to graduation differs across these groups rather than how their endline knowledge differs on average. The covariate coefficients and the remaining selection parameters are reported in Appendix Table D1.

The excluded variable shifts participation in the expected direction, but weakly. The coefficient on distance is -0.032, with a delta-method standard error of 0.016 and a bootstrap standard error of 0.037, so families living farther from the campus are less likely to graduate. A one-standard-deviation increase in distance, 3.0 kilometers in the offered arm of the estimation sample, lowers the graduation index by roughly ten percent of a standard deviation. The effect is modest, consistent with a population living close to campus on average. It is borderline significant under the delta method, with a $z$-statistic of -1.99, and not significant under the bootstrap, and the linear first-stage $F$ is 3.1.[16] The scale of the graduation shock itself is estimated at $\widehat{\sigma}_\omega$ = 0.41 (Appendix Table D1).

The investment equation delivers two of the quantities on which the investment channel depends. The first is inertia, which is strong, and which enters the channel through the marginal product of investment, since $\beta_{X_1}$ is obtained by dividing by it in (18). The coefficient on baseline

[15]The estimation sample is smaller than the endline experimental sample ($N$ = 890) because of missing values in baseline or endline parental investment measures and in the home-school driving-distance shifter. Treatment-control balance in this sample is reported in the last set of columns of Table B3.

[16]Section 6.2 and Appendix F.1 discuss the weakness of distance as a shifter of graduation.

Table 4: Structural Model Estimates

| | | (1) No $G_i \times W_i$ interactions | | | (2) With $G_i \times W_i$ interactions | | |
|---|---|---|---|---|---|---|---|
| Parameter | Description | Estimate | SE (delta) | SE (boot.) | Estimate | SE (delta) | SE (boot.) |
| *Panel A: Graduation equation* | | | | | | | |
| $a_0$ | Intercept (composite) | 0.957 | (1.101) | (1.180) | 1.160 | (1.112) | (1.203) |
| $a_{X_0}$ | Baseline investment $X_{i0}$ | 0.721 | (0.165) | (0.179) | 0.690 | (0.164) | (0.179) |
| $a_{Z_G,\text{distance}}$ | Home-school distance | -0.032 | (0.016) | (0.037) | -0.031 | (0.016) | (0.038) |
| *Panel B: Investment equation* | | | | | | | |
| $\widetilde{\gamma}_0$ | Intercept (composite) | 0.973 | (0.509) | (0.504) | 0.972 | (0.509) | (0.504) |
| $\gamma_{X_0}$ | Baseline investment $X_{i0}$ | 0.809 | (0.097) | (0.098) | 0.809 | (0.097) | (0.098) |
| $\gamma_G$ | Graduation $G_i$ | 0.362 | (0.065) | (0.065) | 0.364 | (0.065) | (0.065) |
| $\rho_\nu$ | Inverse Mills ratio | 0.024 | (0.065) | (0.062) | 0.020 | (0.063) | (0.062) |
| *Panel C: Knowledge production function* | | | | | | | |
| $\beta_0$ | Intercept | -3.031 | (10.149) | (10.786) | -5.479 | (10.063) | (10.622) |
| $\beta_{K_0}$ | Baseline knowledge $K_{i0}$ | 0.358 | (0.038) | (0.038) | 0.340 | (0.047) | (0.049) |
| $\beta_{X_1}$ | Investment $X_{i1}$ | 6.123 | (2.619) | (2.708) | 5.857 | (2.549) | (2.594) |
| $\beta_G$ | Graduation $G_i$ (direct) | 6.248 | (3.308) | (3.322) | 14.077 | (4.629) | (4.566) |
| $\beta_{W,\,K_0\geq\text{med}}$ | $\times\mathbf{1}\{\text{above-median baseline knowl.}K_{i0}\}$ | – | | | -5.801 | (2.298) | (2.370) |
| $\beta_{W,\,\text{educ}\geq\text{med}}$ | $\times\mathbf{1}\{\text{above-median parental schooling}\}$ | – | | | -6.267 | (2.463) | (2.471) |
| $\rho_0$ | Inverse Mills ratio, $G_i = 0$ | -1.240 | (1.615) | (1.635) | -1.473 | (1.601) | (1.622) |
| $\rho_1$ | Inverse Mills ratio, $G_i = 1$ | 1.643 | (7.015) | (6.942) | 0.096 | (7.321) | (7.110) |
| $\rho_1 - \rho_0$ | Selection wedge | 2.883 | (7.005) | (7.050) | 1.568 | (7.290) | (7.125) |

*Note:* Structural parameters of the model of Section 5. Test scores are in percent of items answered correctly, so the Panel C coefficients are in percentage points; investment is measured in weekly reading hours, the scale of the investment factor (Appendix B.4). Column (1) rules out observed heterogeneity in the return to graduation; column (2) allows it to vary with indicators for above-median baseline knowledge and above-median parental schooling. The variable excluded from the investment and knowledge equations is home-to-campus driving distance in kilometers. Both columns condition on the controls of Section 5.3 and on an indicator equal to one when at least one control is missing; the control coefficients, the remaining selection parameters and the scale of the graduation shock appear in Appendix Table D1. Standard errors are from the delta method, with reduced-form standard errors clustered by cohort and school, and from a school-by-cohort cluster bootstrap of the full estimation procedure (2,000 replications).

investment is $\widehat{\gamma}_{X_0} = 0.81$ with a standard error of 0.097, so a family's prior routine strongly predicts its investment in the following year. The second is the response to the program, with graduation raising investment by $\widehat{\gamma}_G = 0.36$ points of the investment factor and a standard error of 0.065. This is larger than the intention-to-treat effect of 0.26 points in Section 4, since $\gamma_G$ is the effect of graduating rather than of being offered a place. Both estimates are essentially unchanged across the two columns, as the model implies, since the heterogeneity terms reach the investment equation only through the control function.

The production function estimates are reported in Panel C. Lagged knowledge carries a coefficient of 0.34 to 0.36. The marginal product of investment is $\widehat{\beta}_{X_1} = 6.12$ in column (1), so one point of the investment factor raises the end-of-year score by about six percentage points. The direct effect of graduation is $\widehat{\beta}_G = 6.25$ percentage points, significant at the ten percent level.

Column (2) enters the two above-median indicators both as controls and in interaction with graduation, which separates two things the earlier specification held together. The controls measure how far apart the groups end the year without graduating. Children with above-median

Table 5: Treatment-Effect Estimates

| | | (1) No $G_i \times W_i$ interactions | | | (2) With $G_i \times W_i$ interactions | | |
|---|---|---|---|---|---|---|---|
| Parameter | Description | Estimate | SE (delta) | SE (boot.) | Estimate | SE (delta) | SE (boot.) |
| ATE | Average treatment effect | 8.463 | (2.789) | (2.701) | 9.133 | (2.924) | (2.791) |
| | *Direct channel* | 6.248 | (3.308) | (3.322) | 7.002 | (3.425) | (3.368) |
| | *Indirect channel* | 2.215 | (1.048) | (1.112) | 2.132 | (1.017) | (1.069) |
| ATT | Average treatment effect on the treated | 9.419 | (1.412) | (1.446) | 9.377 | (1.358) | (1.430) |
| | *Direct channel* | 6.248 | (3.308) | (3.322) | 6.727 | (3.394) | (3.323) |
| | *Indirect channel* | 2.215 | (1.048) | (1.112) | 2.132 | (1.017) | (1.069) |
| | *Selection on unobserved gains* | 0.956 | (2.323) | (2.273) | 0.519 | (2.410) | (2.279) |
| ATU | Average treatment effect on the untreated | 4.897 | (11.161) | (10.974) | 7.627 | (11.752) | (11.278) |
| ITT | Intention-to-treat effect | 7.427 | (1.113) | (1.100) | 7.394 | (1.071) | (1.088) |

*Note:* Effects of graduating from JumpStart, in percentage points of the endline JumpStart score, computed from the estimates of Table 4 via (22)–(24). The ATE is the sum of the direct channel, $\beta_G + \boldsymbol{\beta}_W' \mathbb{E}[\boldsymbol{W}_i]$, and the indirect channel, $\beta_{X_1}\gamma_G$. The ATT evaluates the direct channel at the graduates' means of $\boldsymbol{W}_i$ and adds selection on unobserved gains, and the ATU is built in the same way for non-graduates. The ITT is the graduation rate among offered families times the ATT (Appendix E.3). Standard errors are from the delta method, with reduced-form standard errors clustered by cohort and school, and from a school-by-cohort cluster bootstrap of the full estimation procedure (2,000 replications).

baseline knowledge end the year 3.50 points higher on average, and those with above-median parental schooling 3.07 points higher, each estimated with a standard error of about 2.2 (Appendix Table D1). The interactions measure how the return to graduating differs across the same groups, and both are negative and significant, with above-median baseline knowledge reducing the direct effect by 5.80 points and above-median parental schooling by 6.27. Against a direct effect of 14.08 points for a family below the median on both dimensions, a family above the median on one gains about 8 points, and a family above the median on both gains about 2. The program's direct benefit is concentrated among the least advantaged families, which reproduces within the model the heterogeneity documented in Section 4.

Table 5 translates these estimates into treatment effects. Graduation raises end-of-year knowledge by 8.46 percentage points on average when the return does not vary with observed characteristics, and by 9.13 points when it does, both significant at the one percent level. The decomposition assigns the larger share to the direct channel, 6.25 points in column (1) and 7.00 in column (2), or roughly three-quarters of the total. The investment channel accounts for the remainder, about a quarter of the effect in both columns; it is the product of the two estimates discussed above, and is significant at the five percent level in both columns. The program's own contact with the household is thus the larger part, but the additional home investment it sets in motion is a real and measurable part of the total. The effect on the treated splits the same way, with a third, small and imprecisely estimated term for selection on unobserved gains. Its indirect channel is unchanged, since the investment response and its return are common to all families, and accounts for 23.5 percent of the effect on the treated in column (1) and 22.7 percent in column (2), against 26.2 and

23.3 percent of the ATE.

The effect on the treated is 9.42 percentage points in column (1) and 9.38 in column (2), estimated precisely in both, and about six tenths above the design-based benchmark for the same 842 families, the Bloom ratio of 8.80 in Appendix Table E1, a gap of under half a standard error of either estimate. The effect on the untreated is 4.90 percentage points in column (1) and 7.63 in column (2), but it is estimated with standard errors of about eleven percentage points, so neither figure is distinguishable from the effect on the treated. Two features of the estimates account for this pattern. First, we detect no sorting into graduation on unobserved gains, in that the selection wedge (25) is 2.88 (s.e. 7.01) in column (1) and 1.57 (s.e. 7.29) in column (2), with neither $\rho_0$ nor $\rho_1$ individually distinguishable from zero.

Because these parameters are weakly identified, however, this result reflects the limits of what the data can detect; it is not evidence that such sorting is absent. Second, the gap narrows in column (2) mainly because the estimated wedge falls, from 2.88 to 1.57, once the return to graduation varies with $\boldsymbol{W}_i$. Composition works the other way, since graduates enter with higher baseline knowledge and so have a lower direct return, which raises the effect on the untreated relative to the effect on the treated. The question of selection on unobserved gains therefore remains open.

Before turning to what these estimates owe to the model's structure, we report one internal check. The model-implied intention-to-treat effect is 7.43 percentage points against the 6.94 the experiment delivers on the same 842 families with the regression adjustment of the bounds, a gap of 0.49 that mirrors the one in the effect on the treated and sits well inside the sampling error of either estimate (Appendix Table E1).

## 6 Experimental Bounds and Structural Extrapolations

The experiment identifies the effect of graduation on the families who would complete the program when offered; everything else in Table 5 is produced jointly by the data and the model's structure. This section separates the two contributions. Section 6.1 derives what the average effect would be with no model at all. Section 6.2 presents the marginal treatment effect curve the model implies, and then asks what the data alone say about that curve. The decomposition of the effect into its direct and investment channels has its own non-structural counterpart, presented in Section 6.3.

### 6.1 Bounds on the Average Treatment Effect

The average treatment effect reported in Table 5 is not identified by the experiment alone. Families in the control group cannot graduate, so the data never reveal what a family that declines the program would have achieved had it completed it. The structural model closes this gap parametrically, through a linear technology, jointly normal unobservables, and a control-function correction for selection into graduation. This subsection asks what the data deliver without that structure, and answers the question by constructing nonparametric bounds on the average effect in the tradition of Manski (1997) and Manski and Pepper (2000).

Let $K_{i1}(g)$ denote the potential outcome under graduation status $g \in \{0,1\}$, so that the observed outcome is $K_{i1} = G_i K_{i1}(1) + (1-G_i)K_{i1}(0)$.[17] Every bound below follows a common construction. By the definition of the average treatment effect we write

$$\text{ATE} = \mathbb{E}[K_{i1}(1) - K_{i1}(0)] = \mathbb{E}[K_{i1}(1)] - \mathbb{E}[K_{i1}(0)], \tag{27}$$

and by the law of iterated expectations, each potential outcome mean can be written as a weighted average of subpopulation means

$$\begin{aligned}\mathbb{E}[K_{i1}(g)] &= P \cdot \mathbb{E}[K_{i1}(g) \mid T_i = 1] + (1-P) \cdot \mathbb{E}[K_{i1}(g) \mid T_i = 0] \\ &= P \cdot \Big\{\pi_1 \cdot \mathbb{E}[K_{i1}(g) \mid G_i^* > 0, T_i = 1] + (1-\pi_1) \cdot \mathbb{E}[K_{i1}(g) \mid G_i^* < 0, T_i = 1]\Big\} \\ &\quad + (1-P) \cdot \Big\{\pi_0 \cdot \mathbb{E}[K_{i1}(g) \mid G_i^* > 0, T_i = 0] + (1-\pi_0) \cdot \mathbb{E}[K_{i1}(g) \mid G_i^* < 0, T_i = 0]\Big\} \end{aligned} \tag{28}$$

where $\mathbb{E}[K_{i1}(g) \mid G_i^*, T_i = t]$ is the expected outcome under graduation status $g \in \{0,1\}$ for a family with net gain from graduation $G_i^*$ and treatment offer $t \in \{0,1\}$. Equations (28) partition the population twice. The first partition is by assignment, a share $P$ to the treatment arm and the remaining $1-P$ to the control group. The second is by the latent willingness to graduate, within each arm a share $\pi_t \equiv \Pr(G_i^* > 0 \mid T_i = t)$ who would graduate if offered the program and the remaining $1-\pi_t$ who would not. We keep the subscript on $\pi_t$ until the first assumption below, which equates the two shares and lets us drop it.[18]

One restriction is in place before any of the assumptions below. Indexing the potential outcome by graduation status alone, as $K_{i1}(g)$, asserts that an offer which does not end in graduation leaves endline knowledge unchanged. This exclusion does not follow from randomization, and

[17] One-sided compliance (2) implies the equivalent definition $K_{i1} = T_i \cdot [G_i K_{i1}(1) + (1-G_i)K_{i1}(0)] + (1-T_i) \cdot K_{i1}(0)$.

[18] The bounds below hold without conditioning on covariates, since randomization makes their unconditional versions valid. We condition on the baseline score and the demographic controls only to gain precision, and the figures quoted in the text come from the regression-adjusted columns of Table 6, which differ little from the unadjusted ones.

Table 6: Nonparametric Bounds on the Average Treatment Effect

| | Unadjusted | | Regression adjusted | |
|---|---|---|---|---|
| | Bounds | 95% CI | Bounds | 95% CI |
| Worst case | [–47.90, 52.10] | [–49.31, 53.51] | [–48.74, 51.26] | [–50.10, 52.63] |
| Random assignment (RA) | [–6.87, 14.89] | [–9.72, 17.44] | [–7.82, 13.95] | [–10.57, 16.16] |
| Monotone treatment response (MTR) | [0.00, 52.10] | [0.00, 53.51] | [0.00, 51.26] | [0.00, 52.63] |
| RA + MTR | [7.41, 14.89] | [5.27, 17.44] | [7.08, 13.95] | [5.32, 16.16] |
| ITT estimate | 7.41 | [4.69, 10.12] | 7.08 | [4.83, 9.32] |
| ATT estimate | 9.47 | [5.99, 12.94] | 9.05 | [6.17, 11.92] |

*Note:* Nonparametric bounds on the average effect of graduating from JumpStart on the endline JumpStart score, in percentage points, following Manski (1997) and Manski and Pepper (2000). The worst case uses only the logical range of the score [0, 100]; RA = random assignment and MTR = monotone treatment response. The regression-adjusted columns net out the control variables used to estimate the ITT. The last two rows show the ITT and ATT estimates. Estimates use the endline experiment sample ($N$ = 890). The second and fourth columns show the 95% confidence interval for the parameter of Imbens and Manski (2004), computed from a school-by-cohort cluster bootstrap, and, in the last two rows, the 95% confidence interval for the ITT and the ATT, with standard errors clustered at the cohort-school level.

in our setting it is a genuine restriction rather than a convenience of notation. Both the structure of JumpStart and the observed relationship between attendance and endline knowledge support it, and we maintain it throughout.[19] In (28) we then replace the quantities the data identify by their sample analogues, and the quantities they do not identify by the most extreme values the maintained assumptions allow. On top of it we impose two assumptions, random assignment (RA) and monotone treatment response (MTR). The lottery delivers RA, while MTR restricts the outcome itself.[20] The resulting bounds are reported in Table 6 with a factorial structure, first with no assumptions at all, then with each assumption on its own, and finally with the two combined, so that the marginal contribution of each is visible. Appendix E states each restriction formally, derives every bound, and reports the tests.

The first case we consider is the worst case, which uses only the quantities directly identified

[19] Each JumpStart session builds on the previous one, and materials are handed out session by session. A family with three consecutive unexcused absences is removed from the program and must return them, so leaving early does not give a family a smaller, self-contained dose of the program. Most non-graduates leave early, with 58 percent attending five or fewer of the twenty-two sessions, and among treated non-graduates endline knowledge does not rise with the number of sessions attended (Appendix Table E2).

[20] Assumptions about how families select into graduation, such as the requirement that those more willing to graduate would also score higher without the program (e.g., the monotone treatment selection assumption, see Manski and Pepper, 2000), would tighten these bounds further. We do not impose them because they restrict the one margin a design with one-sided compliance cannot observe, namely who among the controls would have graduated. Appendix E.5 reports the tests of these restrictions and shows that the one contrast isolating selection is estimated too imprecisely to either support or contradict them.

from the observed data, namely the assignment probability $P$, the graduation probability in the treatment arm $\pi$, the control-group mean $\mathbb{E}[K_{i1} \mid T_i = 0]$, and the mean outcomes of graduates and non-graduates among treated families, $\mathbb{E}[K_{i1} \mid T_i = 1, G_i = 1]$ and $\mathbb{E}[K_{i1} \mid T_i = 1, G_i = 0]$.[21] All remaining unidentified means are assigned to the endpoints of the score range, 0 or 100, so as to minimize or maximize the average treatment effect.

The first row of Table 6 reports the resulting bounds. With no restrictions beyond the logical range of the score, they are largely uninformative, spanning -48.7 to 51.3 percentage points and leaving the sign of the effect undetermined.

Additional restrictions on the potential outcomes sharpen these bounds considerably. The assumption we consider first is RA, which requires that the offer be independent of the potential outcomes and of the latent willingness to graduate,

$$\left(K_{i1}(0), K_{i1}(1), G_i^*\right) \perp T_i. \tag{29}$$

Two things follow. The two shares in (28) coincide, $\pi_1 = \pi_0 \equiv \pi$, and assignment carries no information about either potential outcome once the latent willingness to graduate is held fixed. The lottery guarantees both.[22]

RA point-identifies two population moments. The first is the population mean of endline knowledge without graduation, $\mathbb{E}[K_{i1}(0)]$, the second term in (27), which by randomization and one-sided compliance is identified by the control mean $\mathbb{E}[K_{i1}(0)] = \mathbb{E}[K_{i1} \mid T_i = 0]$. The second is the average effect on the treated, which follows directly from the intention-to-treat effect. If treatment assignment has no direct effect on $K_{i1}(g)$, the expected outcome of never-takers is the same in both treatment arms. Their contribution therefore cancels from the ITT, leaving the ATT scaled by the graduation rate $\pi$ (Bloom, 1984).[23] Appendix Table E1 reports the resulting estimates of the effect on the treated, under each combination of sample, regression adjustment and denominator. The second row of Table 6 therefore yields an identified set from -7.8 to 14.0, roughly one-fifth as wide as the worst-case interval.[24]

[21] Under one-sided compliance, treated graduates correspond to families with $G_i^* > 0$ and treated non-graduates to families with $G_i^* < 0$. The control-group mean is a mixture of these two latent groups, while outcomes for control-group graduates are never observed.

[22] RA is the potential-outcome counterpart of Assumption 1 of Section 5.4, which implies it, since the potential outcomes and $G_i^*$ are functions of the model's shocks. The bounds ask for less, since they need no model and use (29) only through the mean independence it implies, equation (44) in Appendix E.2.

[23] See (26) and Appendix E.3 for the derivation.

[24] The only remaining unidentified moment is $\mathbb{E}[K_{i1}(1) \mid G_i^* < 0, T_i = 1]$, the outcome that families who would decline the program would have obtained had they graduated. Because these families account for a share $1 - \pi$ of the population, this moment enters the average effect with weight $1 - \pi$. The RA interval therefore has width $(1 - \pi)$ times the range of the outcome, or about 22 points in our setting, compared with the full 100-point range under the worst case. Appendix E.6 provides the derivation.

The second assumption we consider is MTR, which requires that graduation weakly improves a child's outcome, $K_{i1}(1) \geq K_{i1}(0)$ for every $i$, thereby ruling out harmful effects of completing JumpStart. This assumption concerns the two potential outcomes of the same child and therefore cannot be tested directly by comparing observed treatment and control groups. Its identifying content is limited to ruling out negative treatment effects, so the lower bound becomes zero while the worst-case upper bound remains unchanged, as reported in the third row of Table 6.

Combined with RA, MTR acquires an implication that is directly testable. RA point-identifies the effect on the treated, and MTR requires that effect, and therefore the intention-to-treat effect as well, to be non-negative. The intention-to-treat effect of Section 4.2 is positive and significant, so this necessary implication holds by a wide margin. The fourth row of Table 6 shows the resulting bounds. Under RA and MTR combined, the average effect of graduation lies between 7.1 and 14.0 percentage points.[25] The interval is narrow relative to what is assumed to obtain it, since it rests on nothing stronger than random assignment, the exclusion of the offer, and the requirement that graduating never lowers a child's score.

These bounds also provide a useful check on the structural model of Section 5, since the two rest on different features of the data. The model parameters are estimated from the investment and knowledge equations in Section 5.4, which use the treatment-control difference in mean scores only indirectly, so the average effect the model implies is not fitted to that difference directly. Comparing the model-implied average effect with a set constructed without its parametric assumptions thus provides an informative check on those assumptions. The model passes this check. Its implied average effects, 8.46 and 9.13 percentage points, both lie within our preferred bounds.[26]

The two exercises therefore play distinct but complementary roles. The conclusion that the program has a positive average effect between 7.1 and 14.0 percentage points follows from randomization together with the exclusion of the offer and MTR, whose observable implications are not rejected by the data. These are offer-margin bounds, since the only exogenous variation behind them is the randomized offer. The parametric structure then selects a point within this range, and Section 6.2 assesses that additional identifying content using variation at the graduation margin, the variation in who completes the program among the families offered a seat.

[25] RA and MTR are complementary in bounding the average effect, since neither alone raises the lower bound above zero, which stays at -7.8 under RA and at zero under MTR, while together they raise it to the intention-to-treat effect (Appendix E.6). The lower bound differs slightly from the 7.6 of Table 3 because the conditional means of the bounds carry no strata fixed effects (Appendix Table E1).

[26] This conclusion is also robust to using the model estimation sample. Recomputing the bounds for the 842 families used to estimate the model, rather than the 890 families in the experimental sample, shifts the preferred interval to 6.9 to 13.6 percentage points and leaves both estimates within it.

### 6.2 Marginal Treatment Effects: Model Estimates and Nonparametric Bounds

As established in Section 6.1, the experiment and two assumptions determine the sign of the average treatment effect and bound its magnitude nonparametrically; only the extrapolation to the families who did not graduate depends on the parametric structure. This subsection examines the object that serves as the basis for that extrapolation. The model implies not only the averages of Table 5 but the entire marginal treatment effect (MTE) curve, which gives the average gain at each quantile of the unobserved resistance to graduating (Heckman and Vytlacil, 2005). Whether households sort into graduation on those gains is the question at stake. We show below that these data cannot settle that question, because the experiment randomizes the offer and not graduation itself, so nothing in the design moves a household along the graduation margin while holding its characteristics fixed. The model answers the question through the joint normality assumption, which imposes a shape on the MTE curve. The question becomes what supports extrapolating from the effect on the treated (identified by the experiment) to the effects on the untreated and the broader population (supplied by the MTE curve), and how far the moments of the graduation margin can constrain that extrapolation.

From equation (21), the individual effect of graduating is $\Delta_i = \widetilde{\beta}_G + \boldsymbol{\beta}_W' \boldsymbol{W}_i + \varepsilon_{i1}$, and because the unobserved gain $\varepsilon_{i1}$ enters both $\Delta_i$ and the graduation error $\omega_i$, the model is a generalized Roy model in which households may sort into graduation based on unobserved gains. Define the unobserved *resistance to graduation* $U_{G,i} \equiv 1 - \Phi(\omega_i/\sigma_\omega)$, which is uniformly distributed on [0, 1] by construction, so that $G_i = T_i \cdot \mathbf{1}\{U_{G,i} \leq p_i\}$ with propensity score $p_i = \Phi(\boldsymbol{a}'\boldsymbol{Z}_i)$ (Heckman and Vytlacil, 2005). Under the joint normality of Section 5.4, and with $\boldsymbol{\beta}_W = \mathbf{0}$ (column (1) of Tables 4–5), the curve is nonlinear in $u$ but linear in the normal quantile of $u$:

$$\mathrm{MTE}(u) \;\equiv\; \mathbb{E}\big[\Delta_i \mid U_{G,i} = u\big] \;=\; \widetilde{\beta}_G \;-\; (\rho_1 - \rho_0)\,\Phi^{-1}(u). \tag{30}$$

The level of the curve is the ATE and its coefficient on $\Phi^{-1}(u)$ is the negative of the selection wedge, so that sorting on unobserved gains is equivalent to $\rho_1 \neq \rho_0$, a single linear restriction on two coefficients of the knowledge equation. Every treatment-effect parameter of Section 5.7 is a weighted integral of (30), and averaging the implied household-level ITT over the offered arm reproduces the identity $\mathrm{ITT} = \pi \cdot \mathrm{ATT}$.

The implied MTE curve is plotted in blue in Appendix Figure D1 with its 95% confidence interval, and averaged over windows of resistance in Figure 1; Table 4 reports the slope and Appendix Table D3 the tests of a flat curve. The curve declines from about 15.2 percentage points at the lowest quantiles of resistance to about 1.8 at the highest, staying positive throughout. The decline

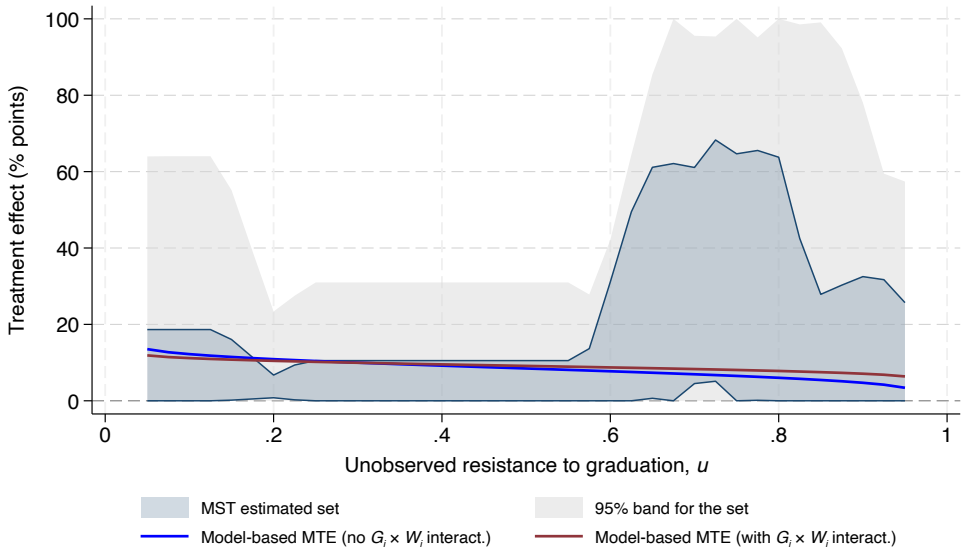
is economically substantial, but the slope is imprecisely estimated, with the wedge $\widehat{\rho}_1 - \widehat{\rho}_0$ equal to 2.88 percentage points (bootstrap s.e. 7.05), and the hypothesis of a flat curve is not rejected ($p$-value 0.68). The implied effect for never-takers, the families a retention policy would newly reach, inherits the imprecision, being positive at the point estimate with an ATU of 4.90 and a delta-method standard error of 11.16 (Table 5). Allowing the gain to vary with observables shifts the level of the curve, $\mathrm{MTE}(u,\boldsymbol{W}_i) = \widetilde{\beta}_G + \boldsymbol{\beta}'_W \boldsymbol{W}_i - (\rho_1 - \rho_0)\,\Phi^{-1}(u)$, with these shifts being precisely estimated as shown in Table 4. Its implied curve computed at the means of $\boldsymbol{W}_i$ is shown in red in the same two figures. The wedge is smaller still in this specification (1.57, bootstrap s.e. 7.13, flat-curve $p$-value 0.83, Table D3), so neither specification can say whether the gain from graduating varies with $u$.

The imprecision of the slope reflects how little of the variation in graduation is driven by cost shifters rather than by household characteristics. The estimated propensity varies little with driving distance, the one regressor excluded from the outcome equations, and mainly with observed household characteristics, above all baseline investment. With the other regressors at their means, a move from the first to the third quartile of distance changes the fitted propensity by only 0.01, against a range of 0.80 when all regressors vary. The fitted propensities also concentrate between 0.61 and 0.93 (the 10th and 90th percentiles), where the normal selection correction is close to linear in the graduation index. Its curvature and the independent variation of distance are what identify the slope of (30), and both are weak in this application (Appendix F.1).

The remaining question is how much evidence supports the model's extrapolation to families who did not graduate. The MTE curves in Figure 1 are nonlinear in $u$ but linear in $\Phi^{-1}(u)$, a shape that follows from the additive separability of the unobserved gain in equation (21) together with joint normality. The experiment identifies the effect on the treated, whereas the effect on the untreated and the exact average treatment effect depend on assuming that the same shape holds over the range of resistance the data do not reach.

We ask what the data alone imply about that shape, using the partial identification framework of Mogstad et al. (2018), hereafter MST. Write the marginal treatment response functions as $m_g(u) \equiv \mathbb{E}\left[K_{i1}(g) \mid U_{G,i} = u\right]$ for $g = 0,1$, so that $\mathrm{MTE}(u) = m_1(u) - m_0(u)$. Every treatment-effect parameter we report is a weighted integral of this pair (Heckman and Vytlacil, 2005), and so is each identified quantity we require the pair to reproduce, which MST call IV-like estimands. MST therefore take every pair $(m_0,m_1)$ that reproduces the chosen estimands and satisfies the shape restrictions we are willing to maintain, and report as its bounds the smallest and largest values the target takes over that set. Within the class of response functions searched over, those two numbers are the most the data and the maintained restrictions can say, since every value

Figure 1: Marginal Treatment Effect Curve and Bounds




*Note:* Bounds following Mogstad et al. (2018) on the MTE averaged over windows of resistance of width 0.10, computed on the model estimation sample ($N$ = 842) with monotone treatment response imposed. The blue and red lines are the model-implied MTE averaged over the same windows, from columns (1) and (2) of Table 5, without and with observed heterogeneity in the return to graduation. The shaded area is a pointwise 95% bootstrap band. Appendix F.3 gives the construction of the windows and of the band.

between them is consistent with both and no value outside them is. The structural model adds additive separability and joint normality and so delivers a single point for each parameter, and comparing that point with the set is a specification check of its functional form.

In our application the targets are averages of the MTE over narrow windows of resistance, which trace the shape of the curve, and the averages that integrate it over wider ranges, the effect on the untreated and the average treatment effect. Nine identified moments constrain the pair. The first is the mean endline score in the control arm, which is $\mathbb{E}[K_{i1}(0)]$ under random assignment. The second is the experimental ITT, the unadjusted difference in mean endline scores between the arms, which carries the restriction $\text{ITT} = \pi\,\text{ATT}$ that one-sided compliance delivers. Six more are the mean score of non-graduates and the increment for graduates within each of three terciles of the fitted graduation propensity among offered families, and the last is the coefficient on distance in a regression of endline knowledge on distance and the demographic controls among offered families.[27] We report two sets of restrictions, the outcome range alone, $0 \leq m_g(u) \leq 100$, and that range together with monotone treatment response, $m_1(u) \geq m_0(u)$ at every $u$. The graduation propensity is the model's probit, and baseline knowledge and the demographic controls enter $m_0$

[27] The effect on the treated is also held at the value the experiment identifies, 8.76 percentage points.

Table 7: Nonparametric Bounds on the Graduation Treatment Effects

| Shape restrictions | ATE | ATU |
|---|---|---|
| Outcome bounds only | [3.09, 10.45] | [-19.09, 17.07] |
| *95% CI for the parameter (Imbens–Manski)* | [-3.33, 17.27] | [-47.84, 49.33] |
| Monotone treatment response (MTR) | [8.17, 10.20] | [5.87, 15.84] |
| *95% CI for the parameter (Imbens–Manski)* | [5.74, 13.91] | [2.54, 25.70] |
| Model estimates: no $G_i \times W_i$ interactions | 8.46 | 4.90 |
| Model estimates: with $G_i \times W_i$ interactions | 9.13 | 7.63 |

*Note:* Nonparametric bounds on the ATE and the ATU, in percentage points, following Mogstad et al. (2018), on the model estimation sample ($N$ = 842). *Outcome bounds only* restricts the marginal treatment response functions to the range of the score, $0 \leq m_g(u) \leq 100$ for $g = 0,1$. *Monotone treatment response* adds $m_1(u) \geq m_0(u)$ for all $u$. Both hold the effect on the treated at the value the experiment identifies, 8.76 (Appendix Table E1). The last two rows report the model estimates of Table 5, columns (1) and (2), for comparison. Intervals are 95% confidence intervals for the parameter of Imbens and Manski (2004), from a school-by-cohort cluster bootstrap, and should be read as indicative (see Appendix F.3).

and $m_1$ with the same coefficients, so that they shift the level of knowledge without shifting the gain from graduating.[28]

Figure 1 plots the model's curve against the sets under MTR, and Appendix Figure F3 compares the sets with and without the restriction. The model lies inside the set almost everywhere. It falls outside only around $u$ = 0.2, where the model puts the average gain above ten percentage points and the data cap it at 6.7 to 9.4. Above $u$ = 0.6 the sets widen sharply, with upper bounds above 60 percentage points between 0.65 and 0.80. The moments constrain the average gain within each tercile of the propensity but not how it varies inside the tercile, so over the range of resistance where the fitted propensities lie the data leave the shape of the curve largely free. The data therefore check the model where they bind, and elsewhere the curve is the model's alone, so the exercise neither rejects the normal specification nor corroborates it.

Table 7 reports the bounds on the average treatment effect and on the effect on the untreated, first requiring only that the two response functions stay within the range of the test score and then adding MTR, with 95% confidence intervals for the parameters. Under the score range alone the average effect lies in [3.1, 10.5]. Adding MTR raises the lower bound to 8.2, well above the zero the restriction alone would impose, and narrows the set to [8.2, 10.2], which contains both model

[28] Appendix F.3 sets out the estimation in full, including the construction of the moments, the shape restrictions, and the confidence intervals.

estimates, 8.46 without observed heterogeneity and 9.13 with it. For the effect on the untreated the set under MTR is [5.9, 15.8]. It contains the estimate with heterogeneity, 7.63, but not the estimate without it, 4.90, which lies 1.0 point below the lower bound, although inside the 95 percent interval of [2.5, 25.7]. The graduation-margin moments therefore favor the specification in which the gain from graduating varies with observed characteristics.

Comparing these sets with the offer-margin bounds of Section 6.1 puts the graduation margin's contribution in perspective. For the average effect the graduation-margin moments sharpen the offer-margin result, since the set under MTR, [8.2, 10.2], lies inside the [6.9, 13.6] that the offer margin gives on the same 842 families (Appendix Table E3), and inside the [7.1, 14.0] of the full experiment sample (Table 6). The two sets rest on different information and different assumptions. The graduation-margin bounds use more identified moments, which describe how outcomes move with the propensity to graduate, and they also restrict how covariates enter the response functions, so the narrower set is not obtained from the offer-margin assumptions alone (Appendix F.3). The effect on the untreated is sharpened much further. On the offer margin it lies in [0.0, 31.4] on the same 842 families and in [0.0, 31.6] on the full experiment sample, bounds that follow from those on the average effect (Appendix E.3). The graduation-margin moments narrow it to [5.9, 15.8], since the small never-taker share multiplies each point of width in the average effect by nearly five. The set could have fallen outside, and the maintained assumptions would have been contradicted. Instead the one parameter only this margin can restrict is sharpened, to a range that contains the model estimate with observed heterogeneity and narrowly excludes the estimate without it.

Taken together, the two subsections establish precisely what each finding rests on. First, that the program is effective for participating families, and the exact magnitude of that gain, is an empirical result, since the treatment effect on the treated follows from random assignment, one-sided compliance, and the exclusion that the offer matters only through completion. Second, establishing that the average effect across all offered households lies between 7.1 and 14.0 percentage points requires only the experimental design and a monotone treatment response assumption, a restriction whose necessary observable implications are not rejected by the experimental data, and the graduation-margin moments, under further restrictions on the response functions, narrow that range to 8.2 to 10.2. Third, whether households sort into graduation based on unobserved gains is an open question that our data cannot resolve.

The effect on the untreated and the point estimate of the average treatment effect are the model's own contribution, obtained from joint normality where the data run out. Both were checked against bounds they could have failed, and only the specification without observed heterogeneity

fails, narrowly, for the effect on the untreated. Whether the decomposition into direct and investment mechanisms depends on the model in the same way is the subject of the next subsection.

### 6.3 The Decomposition Under Weaker Assumptions

The decomposition based on the model's estimates in Section 5 assigns roughly a quarter of the effect of graduation to the indirect channel (both on average and among graduates) and the remaining three quarters to the direct channel. The direct channel is the effect of graduation at unchanged measured investment, which includes the liaison's monthly sessions with the child and any input the investment index does not record. The indirect channel is the increase in knowledge produced by the additional parental investment that graduation induces. The most closely related study reaches the opposite conclusion. Attanasio et al. (2020) find that the cognitive gains from the Colombian home-visiting intervention are explained entirely by the increase in parental investment, with a direct effect that is small and statistically indistinguishable from zero. The contrast raises a natural question, namely how much of our decomposition is produced by our structural model. This subsection answers it in two ways. First, we reproduce the decomposition's qualitative conclusion with a mediation analysis that dispenses with some of the model's parametric structure. Identification rests on a single exclusion restriction, that baseline investment affects endline knowledge only through its effect on endline investment, so that it can serve as an instrument for the latter. The structural model separates the two channels using this same restriction, so what changes in this subsection is the apparatus built around it, not the identifying content itself. Second, we relax the exclusion by allowing baseline investment to affect endline knowledge directly, and ask how strong this dependence would have to be to overturn the decomposition. Finally, we return to the contrast with Attanasio et al. (2020) and ask whether the two studies' findings need reconciling at all.

**The mediation analysis.** We implement the decomposition with fewer of the parametric restrictions imposed by the model in Section 5, using the following system of equations:

$$X_{i1} = \gamma_0 + \gamma_1 T_i + \gamma_2 X_{i0} + \boldsymbol{\gamma}_3' \boldsymbol{Z}_{1,i} + u_i \tag{31}$$

$$K_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 T_i + \boldsymbol{\beta}_3' \boldsymbol{Z}_{2,i} + v_i. \tag{32}$$

These are the mediation-analysis counterparts of the investment and knowledge equations (7) and (3), and their coefficients are the reduced-form counterparts of the model's $\gamma$ and $\beta$, not the same parameters. Equation (31) writes endline investment $X_{i1}$ as a linear function of treatment

Table 8: Mediation Analysis: Decomposition of the Assignment Effect (Pooled Sample)

| | Return to investment ($\widehat{\beta}_1$) | Direct effect ($\widehat{\beta}_2$) | Effect on investment ($\widehat{\gamma}_1$) | Total effect | Mediated share (%) |
|---|---|---|---|---|---|
| OLS (Bartlett factor score) | 2.126 | 6.396*** | 0.285 | 7.003 | 8.7 |
| | (0.622) | (1.168) | (0.053) | (1.123) | (3.5) |
| OLS (regression factor score) | 7.438 | 6.334*** | 0.090 | 7.003 | 9.6 |
| | (2.076) | (1.171) | (0.016) | (1.122) | (3.8) |
| OLS (regression score return, Bartlett response) | 7.438 | 4.881*** | 0.285 | 7.003 | 30.3 |
| | (2.076) | (1.366) | (0.053) | (1.123) | (12.0) |
| Instrumental variables (IV) | 6.903 | 5.048*** | 0.285 | 7.018 | 28.1 |
| | (2.449) | (1.361) | (0.053) | (1.117) | (12.5) |
| *AR 95% set for* $\beta_1$ | [2.197, 12.493] | | | | |

*Note:* The table decomposes the assignment effect on endline knowledge using equations (31) and (32), estimated on the model estimation sample, $N = 842$. The columns report the return to endline investment, the direct effect of assignment, the effect of assignment on endline investment, the total effect $\hat{\beta}_1\hat{\gamma}_1 + \hat{\beta}_2$, in percentage points, and the mediated share $\hat{\beta}_1\hat{\gamma}_1/(\hat{\beta}_1\hat{\gamma}_1 + \hat{\beta}_2)$, in percent. The first two rows are ordinary least squares, measuring endline investment by the Bartlett and by the regression factor score respectively; the third combines the return to investment of the second row with the effect of assignment on the Bartlett score of the first, and obtains the direct effect as the total effect minus the indirect one; the fourth instruments endline investment with baseline investment, and the row below it reports, in the first column, the 95% Anderson–Rubin confidence set for $\beta_1$. The first-stage $F$ of the instrumental-variables row, the Kleibergen–Paap rk Wald statistic, is 55.9. Controls are those of the knowledge equation. Standard errors are in parentheses, from a school×cohort cluster bootstrap of the whole system (500 replications) for the mediated share, the total effect and the direct effect of the third row, and clustered by school×cohort for the other coefficients. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Stars are shown on the direct effect only.

assignment $T_i$ and baseline investment $X_{i0}$. Equation (32) writes endline knowledge $K_{i1}$ as a linear function of endline investment and of assignment itself. The control vectors $\boldsymbol{Z}_{1,i}$ and $\boldsymbol{Z}_{2,i}$ contain the same variables as the corresponding vectors of the structural model, with baseline knowledge $K_{i0}$ stacked inside both, so that in our application the two vectors coincide.

The system separates the effect of being offered the program into the same two components as the structural model. Assignment raises endline investment by $\gamma_1$, and each unit of investment raises endline knowledge by $\beta_1$, so its *indirect effect* is $\beta_1\gamma_1$. The component that does not pass through measured investment is, instead, $\beta_2$, the *direct effect* of the treatment assignment. The total effect is therefore $\beta_1\gamma_1 + \beta_2$, and the mediated share is $\beta_1\gamma_1/(\beta_1\gamma_1 + \beta_2)$.

Table 8 reports the estimates of (31) and (32) and the mediated shares they imply. The least-squares estimates in the first row attribute 8.7 percent of the total effect to measured investment (s.e. 3.5 percentage points) and the remaining 91.3 percent to the direct channel. The direct effect of assignment is 6.4 percentage points against a total effect of 7.0, so assignment reaches knowledge largely without passing through the investment we measure.[29]

Two problems, however, bias the least-squares estimate of the return to investment, on which the decomposition rests. The first is measurement error in the endline investment factor, which biases the estimated return toward zero and so overstates the direct effect. The second is the possible endogeneity of investment, since parents may choose it in response to determinants of

[29] The total effect equals the intention-to-treat effect of assignment on endline knowledge on this sample, conditional on baseline investment and the controls of the knowledge equation, 7.02 percentage points, exactly in the instrumental-variables row and to within 0.02 in the least-squares rows. Divided by the graduation rate of 0.789, it implies an effect on graduates of 8.9 percentage points, in line with the model-sample estimates in Table E1.

endline knowledge that we do not observe. Unlike attenuation, such a bias could run in either direction.

We first ask how much of the bias is measurement error. The second row of Table 8 re-estimates the same system with endline investment measured by the regression factor score, which does not transmit the attenuation the Bartlett score carries when it enters as a regressor.[30] The return to investment rises from 2.13 to 7.44. Combined with the effect of assignment on the Bartlett score, as in the third row, it raises the mediated share to 30.3 percent (s.e. 12.0 percentage points).[31] Correcting measurement error alone therefore brings the mediated share close to the share of the investment channel in the structural model, 26.2 percent of the average treatment effect and 23.5 percent of the effect on graduates (Table 5).[32]

While the regression factor score corrects only for attenuation, instrumenting endline investment can correct for both attenuation and endogeneity. This requires an instrument that predicts true investment but is unrelated to its measurement error and to unobserved determinants of endline knowledge. We use baseline investment as the instrument and estimate (31) and (32) by two-stage least squares. Baseline investment is realized before $T_i$ is randomized, so it is orthogonal to assignment by design, and it removes the attenuation provided its measurement error is uncorrelated with that of endline investment. The remaining requirement is the exclusion restriction already maintained in the model, that baseline investment affects endline knowledge only through the endline investment it predicts. The fourth row of Table 8 reports these estimates. The mediated share rises to 28.1 percent (s.e. 12.5 percentage points), roughly three times the least-squares figure of the first row and close to the 30.3 percent of the third, with a first-stage $F$ of 55.9. Its 95 percent bootstrap confidence interval runs from 7.9 to 55.2 percent, so it excludes full mediation but not a mediated channel larger than the direct one, and the direct channel is the larger of the two at the point estimate but not with statistical confidence.[33] The structural decomposition of Table 5, which applies the same correction through the inertia ratio of equation (18), gives shares of the investment channel statistically indistinguishable from this mediation-based estimate.

We next use variation across observable groups for two complementary checks. The first is a

[30] See footnote 10 and Appendix B.4.

[31] Using the regression score in both equations, as in the second row, does not correct the share, since the shrinkage that raises $\hat{\beta}_1$ lowers $\hat{\gamma}_1$ by about as much (Appendix B.4), and the score's prediction error, which is correlated with the offer, biases $\hat{\beta}_2$. The third row therefore computes the indirect effect as $\hat{\beta}_1$ of the second row times $\hat{\gamma}_1$ of the first, and the direct effect as the total effect of the first row minus this indirect effect.

[32] Under one-sided compliance the mediation share is a share of the effect on the treated, so its exact structural counterpart is the ATT share.

[33] We report the percentile interval from the school×cohort cluster bootstrap, which is asymmetric around the point estimate since the share, as a ratio, has a skewed sampling distribution.

reduced-form proportionality test.[34] If assignment affects knowledge only through measured investment, and the return to investment is common across families, then the assignment effect on knowledge must be proportional to the assignment effect on investment. The ratio of the two intention-to-treat effects is then the same in every group. We assess this restriction through an over-identification test, splitting the sample at the median of a baseline characteristic. The test uses only the variation induced by assignment and does not rely on baseline investment as an instrument.[35] Appendix Table G1 reports the two ratios and the $p$-value of this test for each splitting variable. The restriction is not rejected for any of the four splits, on the baseline JumpStart score (our measure of baseline knowledge), parental schooling, the Bracken baseline score, or the child's age. The closest is parental schooling, with a $p$-value of 0.101, where the ratio is 44.6 below the median and 16.4 above it, so the effect on knowledge per unit of the effect on investment is nearly three times larger in families with less schooling.

The second check brings baseline investment back as an instrument and asks whether the pooled decomposition, and its rejection of full mediation, survive when the direct effect, the return to investment, and the response of investment to assignment are allowed to differ across median splits on the baseline JumpStart score, parental schooling, the Bracken baseline score, and the child's age at pretest (Appendix G.1). Full mediation is rejected in every one of the twelve group-by-specification cells of Appendix Tables G2, G3 and G4, with a largest $p$-value of 0.002, so allowing the parameters to differ across groups does not remove the direct channel. Under the most flexible specification, which lets all three parameters differ across groups, the mediated share ranges from 17.6 to 32.2 percent across the groups defined by the baseline JumpStart score, the Bracken score, and the child's age, and no test of equal returns or equal direct effects rejects for these partitions.

Parental schooling is the exception. When all three parameters are group-specific, the returns do not differ across groups ($p$ = 0.584), while the direct effects do, at 9.7 percentage points below the median and 2.9 above it ($p$ = 0.039; Appendix Table G4). When only the direct effect is group-specific, the contrast is similar, 9.1 percentage points below the median and 3.1 above it, and equality is rejected with a $p$-value of 0.034, although this specification holds the investment response common across groups, while the most flexible specification puts it at 0.265 below the median and 0.306 above it. The pattern is consistent with the education gradient of Section 4.4 op-

[34] We thank an anonymous referee for suggesting this across-subgroup proportionality test.

[35] With the effect of assignment on investment group-specific and the return and the direct effect common to all groups, the ratio of the two intention-to-treat effects in group $h$ is $\mathrm{ITT}_{K,h}/\mathrm{ITT}_{X,h} = \beta_1 + \beta_2/\gamma_{1h}$, which equals $\beta_1$ in every group under full mediation. The test maintains a direct effect common to all groups. Appendix G.1 derives the test and discusses its power.

erating through the direct channel rather than through measured investment, and with the structural model, which places the parental-schooling gradient in the graduation effect rather than in the return to investment.[36]

**Sensitivity to the baseline-investment exclusion.** The instrumental-variables estimate of the return to investment in Table 8 is more than three times the least-squares estimate, a gap that the correction for attenuation in the second row closes. This estimate rests on the exclusion that baseline investment affects endline knowledge only through endline investment, which cannot be tested with a single instrument, so we quantify how the decomposition would change were the exclusion violated.

Suppose that baseline investment enters the knowledge equation directly,

$$K_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 T_i + \delta X_{i0} + \tilde{v}_i, \tag{33}$$

where $\delta$ captures the effect of baseline investment on $K_{i1}$ that does not operate through endline investment $X_{i1}$, and $\tilde{v}_i$ is the error of (32) net of $\delta X_{i0}$. The controls $\boldsymbol{Z}_{2,i}$ are included in estimation and omitted from the notation. Setting $\delta = 0$ recovers the exclusion imposed in (32), and the violation grows with $|\delta|$. A positive $\delta$ corresponds to reinforcing violations, in which families that invest more at baseline also experience larger knowledge gains for reasons other than endline investment. A negative $\delta$, by contrast, reflects compensating violations, in which parents invest more at baseline because their child is behind in ways not captured by baseline knowledge, with that shortfall persisting to endline.

As baseline investment cannot both serve as an instrument and affect endline knowledge on its own, $\delta$ and $\beta_1$ are not jointly identified in the two-stage least squares system estimated above. We can nonetheless ask how the decomposition would change if the exclusion were violated by a given amount $\delta \neq 0$. Following Imai et al. (2010) and Conley et al. (2012), for each value of $\delta$ on a grid spanning both positive and negative values, we re-estimate the same two-stage least squares system on the adjusted outcome $K_{i1} - \delta X_{i0}$.[37] We then report the values of $\delta$ at which each conclusion about the decomposition no longer holds, both in levels and as a fraction of the

[36] With four partitions examined, a Bonferroni correction at the 5 percent level requires $p < 0.0125$. Neither rejection of equal direct effects, with $p = 0.034$ when only the direct effect is group-specific and $p = 0.039$ when all three parameters are, falls below this threshold.

[37] The violation is removed exactly when $\delta$ equals its true value. The adjustment does not change the total effect of assignment, $\beta_1\gamma_1 + \beta_2$, but only how it is split between the direct and indirect channels, since each unit of $\delta$ lowers the return by $1/\hat{\gamma}_2$ and raises the direct effect by $\hat{\gamma}_1/\hat{\gamma}_2$ (Appendix G.2). Selection into graduation that operates through baseline investment is observationally equivalent to a direct effect of $X_{i0}$, so $\delta$ captures that departure as well as a failure of the exclusion.

estimated return $\hat{\beta}_1$, which is expressed in the same unit.[38] These breakdown values measure how large a violation would have to be to overturn each conclusion, so that the reader can judge whether a violation of that size is plausible. We also report 95 percent confidence intervals for the direct effect and the mediated share at each value of $\delta$. Figures G1 and G2 trace both channels along the grid, and Appendix Table G5 reports the breakdown values. Appendix G.2 provides implementation details.

The decomposition responds differently to violations of the two signs. Reinforcing violations raise the direct effect and lower the mediated share. They can therefore remove the mediated channel, which disappears once $\delta$ reaches 5.5 (80 percent of the estimated return), but never the direct one. Compensating violations lower the direct effect, and only large ones overturn it. The direct effect remains significant at the 5 percent level until $\delta$ = -5.4 (78 percent of the estimated return), and it falls to zero, the case of full mediation, only at $\delta$ = -14.2 (2.1 times the estimated return).[39]

To judge whether violations of the sizes reported in Table G5 are plausible, we consider a simple model in which the exclusion fails. Suppose the shock to the investment policy function is persistent, with first-order autocorrelation $\rho_u$, and correlated with the knowledge shock, with slope $\psi$. Let $\gamma_2$ denote structural inertia, i.e., the effect of baseline investment on endline investment that does not operate through the persistent shock. The persistent shock can represent, for instance, a stable home environment, which both leaves parents more time to invest and supports the child's learning directly ($\psi > 0$), or a lasting difficulty of the child, to which parents respond by investing more in both years ($\psi < 0$). Appendix G.3 describes this model in detail. Under this framework, the violation is given by $\delta = \rho_u \psi$, which implies that the return to investment estimand is

$$\operatorname{plim} \hat{\beta}_1 = \beta_1 + \frac{\rho_u \, \psi}{\widetilde{\gamma}_2}, \tag{34}$$

where $\widetilde{\gamma}_2 \equiv \gamma_2 + \rho_u$. The first stage identifies only the sum $\widetilde{\gamma}_2$, estimated at 0.802 with a standard error of 0.107, but not its two components $\gamma_2$ and $\rho_u$.

We restrict attention to $\gamma_2 \geq 0$ and $\rho_u \geq 0$. These restrictions cannot be tested, but they strike us as the more reasonable ones. The first means that a family's baseline investment routine does not by itself lower the investment it targets at endline. The second means that the investment shock persists rather than reverses, so that the unobserved circumstances that raise investment at

[38] Both indices are anchored on weekly reading time, which gives them a common unit under the assumption that reading time relates to latent investment in the same way at baseline and at endline (Appendix B.4).

[39] In standard-deviation units this violation is -0.31, so that a one standard deviation increase in baseline investment, net of the controls, lowers endline knowledge by 0.31 standard deviations of the knowledge-equation error.

baseline, such as the quality of interaction between parent and child, support from other adults, or the home learning environment, also tend to raise it at endline. Together, they imply that neither $\gamma_2$ nor $\rho_u$ can exceed the first-stage coefficient of 0.802, and that the bias in (34) has the sign of $\psi$ (Appendix G.3). When $\psi$ is positive, favorable conditions for learning accompany high investment, so that the return to investment is overstated and the direct effect understated, and our decomposition is conservative about the direct channel. Only a negative $\psi$, as when parents invest more in a child who falls behind, can make the direct effect appear larger than it is.

The last column of Table G5 reports the bias in the return implied by each breakdown value, $\delta/\widetilde{\gamma}_2$, which under the two restrictions is also the smallest $|\psi|$ that can generate it. The first result is that full mediation requires a strongly negative $\psi$, that is, parents who invest much more when their child's learning falls short. The direct effect vanishes at $\delta$ = -14.2, which implies a true return to investment of 24.6 against our estimate of 6.9, and requires $\psi \leq -17.7$. In words, each unit of the investment shock must be accompanied by a knowledge shock at least 17.7 percentage points lower, an association at least 2.6 times as strong as the estimated effect of investment itself. This bound is reached only in the case most favorable to full mediation, in which the whole first stage reflects persistence of the shock ($\rho_u = \widetilde{\gamma}_2$ = 0.802), and any other split of the first stage requires a stronger association. If half of it reflects structural inertia instead ($\rho_u = \gamma_2$ = 0.401), the requirement doubles to $\psi \leq -35$, more than five times the estimated return. In our data, full mediation, as found by Attanasio et al. (2020), therefore requires compensation of this strength.

Our finding that the direct channel is the larger of the two is far less robust. The two channels contribute equally at $\delta$ = -4.33, where the true return would be 12.3, so that our estimate would be biased downward by 5.4. This requires a negative $\psi$ at least 0.78 times as large as the estimated effect of investment, less than a third of what full mediation requires. The result is consistent with the confidence interval for the mediated share reported above, which contains one half, so that the data cannot reject that the two channels contribute equally. In this canonical case, full mediation therefore requires compensation that is hard to credit, while a mediated channel larger than the direct one requires much less.

**Discussion.** A correction of the same kind we perform operates in Attanasio et al. (2020), though with a different instrument set, local prices of investment goods rather than lagged investment. There, instrumenting investment raises the mediated share of the cognitive impact from about 25 percent under ordinary least squares to essentially all of it. In both papers the naive mediation understates the investment channel and the correction moves the split toward mediation; the papers differ in where the correction comes to rest. Our decomposition is the more conservative of the

two. Corrected with and without the model, the investment channel accounts for about a quarter of our program's effect, and the direct channel carries the rest.

The remaining divergence has at least three possible explanations, which a comparison across the two studies cannot separate. The first is measurement. Attanasio et al. (2020) measure material investment through interviewer observation and attribute the null role of self-reported time investment partly to reporting error; every item in our investment factor is parent-reported, and our factor weights the quantity of interaction. Whatever dimensions of parenting a quantity-weighted, self-reported index does not record, improvements in the quality and technique of interaction above all, are absorbed into our direct effect by construction, as Section 5 defines it.

The second explanation is identification. If the return to investment is heterogeneous across families, each instrument identifies a differently weighted average of it (Heckman and Vytlacil, 2005). Prices weight the families whose goods investment responds to prices; lagged investment weights the families whose investment persists, which in our factor means interaction-intensive investment above all. Even with both instruments valid and both mediators perfectly measured, the two corrections could come to rest at different shares because they average different families' returns. The data of either study cannot rule this in or out.

The third explanation is that the programs themselves differ, in the age they serve, in delivery, and in outcome. At the ages JumpStart serves, specific school-readiness content is directly teachable, the curriculum's content is the test's content, and children attend one session with the liaison each month. The module-level pattern is consistent with such a channel: treatment effects concentrate on the most curriculum-specific modules (name recognition and book handling), which a family liaison can convey directly to a child during monthly sessions. By contrast, effects on the curriculum-distant Bracken assessment remain small. While this pattern cannot fully disentangle the liaison's direct instruction from unmeasured, targeted home teaching by parents, it is difficult to reconcile with a generalized increase in parental stimulation, which would not yield such localized gains.

The answer to the question posed at the outset, then, is that the two findings do not need reconciling. The channel composition of parenting programs is not a universal constant: it varies with the population served, the program's design, what is measured as investment, and the method of the decomposition itself, including the identification strategy used to uncover the mechanism. Mechanism findings do not transport from one program to another, which is why a program evaluation that aims at mechanisms must estimate its own decomposition and state what the estimate rests on. Separating the readings within a single study would require investment measurement

rich enough to capture quality and technique alongside quantity, a design point we return to in the conclusion.

# 7 Conclusion

This paper evaluates JumpStart, a low-cost, parent-directed intervention designed and operated by the Alief Independent School District to prepare children for pre-K. Access was randomized across three cohorts with zero control-group crossover. The experimental design identifies the intention-to-treat (ITT) effect directly; with one-sided compliance and the exclusion that the offer matters only through completion, the compliance rate converts it into the average treatment effect on the treated (ATT), the effect on families who complete when offered. The offer of JumpStart increased scores on the district's curriculum-aligned assessment by 7.6 percentage points and boosted parental investment in home learning; its effect on the broader Bracken school-readiness composite is smaller and imprecisely estimated. For families completing the program, the effect on the curriculum-aligned assessment is 9.8 percentage points.

Our mechanism findings warrant particular attention from future evaluations. Roughly one-quarter of the program's effect operates through measured parental investment, with the remainder reaching the child without passing through measured investment. Both the structural model and the mediation analysis yield similar shares after correcting for measurement error and omitted variable bias. The direct channel remains statistically significant under any reinforcing violation of the baseline-investment exclusion, in which families that invest more at baseline also gain more for reasons other than endline investment, and under compensating violations up to 78 percent of the estimated return to investment, although its dominance over the investment channel is less robust than its presence. Nor is this split a universal feature of early intervention. Applying a correction of the same kind to a different program and population, Attanasio et al. (2020) attribute virtually the entire cognitive gain to home stimulation. Mechanism shares are program-specific. A policy maker who funds parenting programs according to their measured effect on home investment would have overlooked much of this program's value.

Policy implications naturally align with the strength of the underlying evidence. Regarding the expansion margin (i.e., offering JumpStart to additional families similar to our sample), the experiment provides direct guidance: the observed impacts were generated at a cost of $450 per family. Regarding the retention margin (i.e., inducing participation among currently non-complying families), the policy-relevant parameter is the effect on the families a given retention policy would bring to completion, which equals the treatment effect on the untreated (ATU) for a policy that

reaches every non-complier. The ATU point estimate is positive but imprecisely estimated, and the data constrain it only through bounds. We therefore frame the expansion implication with confidence and the retention implication conditionally.

Whether these initial gains will persist remains an open question, with two plausible hypotheses pointing in opposite directions. Because the curriculum is intentionally targeted, narrow gains might quickly fade. However, this narrowness directly reflects input from the district's pre-K teachers regarding foundational skills. If those teacher assessments are informative, dynamic complementarities between early skills and subsequent instruction should compound rather than erode these gains over time. Resolving this empirical question requires tracking these cohorts into pre-K classrooms, a task made feasible by our ongoing institutional partnership.

Finally, our findings offer two methodological takeaways for future study design. First, randomized variation in the cost of completing the program, through participation incentives or subsidized access, would identify the effect on the families it induces to complete without parametric structure. Incentives may, however, change the program itself, since families induced by rewards may engage differently from families induced by conviction. The reach of such variation is limited by the cost elasticity of participation, which our distance evidence suggests is low in this population, and by the requirement that the incentives remain small relative to household resources, since payments large enough to move the most resistant families would affect the home environment directly.

Second, because mechanism decomposition depends critically on home inputs, future evaluations of parent-focused interventions should invest in richer, objective measurement systems specifically designed for structural estimation. Our own decomposition indicates why: correcting measurement error in the investment factor closes most of the gap between the naive and the corrected mediated share, so the quality of the investment measure, more than the endogeneity of investment, limits what mechanism analysis can deliver.

Readers may take these results at their preferred level of assumption. The lottery alone identifies a 7.6 percentage point effect of being offered JumpStart. Adding one-sided compliance and the exclusion that the offer matters only through completion (Appendix E.3) implies a 9.8 percentage point effect for families who would complete the program when offered. Adding monotone treatment response, whose refutable implications are not rejected in the data, bounds the average effect across all offered families between 7.1 and 14.0 percentage points, and the moments of the graduation margin, under further restrictions, narrow this range to 8.2 to 10.2. Accepting the parametric selection model delivers the exact average effect, roughly 9 percentage points, the

effect on families who would not complete, and the one-quarter mechanism share through home investment, and our partial-identification analysis shows how much these estimates depend on its functional form. Whether families sort into completion on unobserved gains is a question our data cannot answer, but little in our core conclusions hinges on the answer. Each step is labeled with what it assumes; readers decide where the evidence satisfies their threshold.

## References


**Andrew, Alison, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina**, "Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial," *Journal of Child Psychology and Psychiatry*, 2020, *61* (6), 644–652.

**Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro**, "Transparency in Structural Research," *Journal of Business & Economic Statistics*, 2020, *38* (4), 711–722.

**Attanasio, Orazio P., Camila Fernández, Emla O. A. Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina**, "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial," *BMJ*, 2014, *349*, g5785. Erratum in: BMJ. 2014;349:g6126.

**Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina**, "Estimating the production function for human capital: results from a randomized controlled trial in Colombia," *American Economic Review*, 2020, *110* (1), 48–85.

**Baker, Amy JL, Chaya S Piotrkowski, and Jeanne Brooks-Gunn**, "The effects of the Home Instruction Program for Preschool Youngsters (HIPPY) on children's school performance at the end of the program and one year later," *Early Childhood Research Quarterly*, 1998, *13* (4), 571–588.

**_ , _ , and _** , "The Home Instruction Program for Preschool Youngsters (HIPPY)," *The Future of Children*, 1999, *9* (1), 116–133.

**Bassok, Daphna, Jenna E Finch, RaeHyuck Lee, Sean F Reardon, and Jane Waldfogel**, "Socioeconomic gaps in early childhood experiences: 1998 to 2010," *AERA Open*, 2016, *2* (3), 2332858416653924.

**Becker, Gary S. and Kevin M. Murphy**, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

**Bloom, Howard S.**, "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 1984, *8* (2), 225–246.

**Bond, Timothy N and Kevin Lang**, "The evolution of the Black-White test score gap in Grades K–3: The fragility of results," *Review of Economics and Statistics*, 2013, *95* (5), 1468–1479.

**Bracken, Bruce A**, *Bracken School Readiness Assessment*, third ed., Psychological Corporation, 2007.

**Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi**, "Plausibly Exogenous," *The Review of Economics and Statistics*, 2012, *94* (1), 260–272.

**Constantinides, George M.**, "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 1990, *98* (3), 519–543.

**Corso, Phaedra S., Justin B. Ingels, and Rebecca L. Walcott**, "Costs of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Evaluation," OPRE Report 2022-01, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services 2022.

**Cunha, Flavio and James Heckman**, "The technology of skill formation," *American Economic Review*, 2007, *97* (2), 31–47.

**Duesenberry, James S.**, *Income, Saving, and the Theory of Consumer Behavior*, Cambridge, MA: Harvard University Press, 1949.

**Gertler, Paul, James J. Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor**, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *344* (6187), 998–1001.

**_ , James J Heckman, Rodrigo Pinto, Susan M Chang, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright**, "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31," Working Paper 29292, National Bureau of Economic Research September 2021.

**Grantham-McGregor, Sally M., Christine A. Powell, Susan P. Walker, and John H. Himes**, "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study," *The Lancet*, 1991, *338* (8758), 1–5.

**Heckman, James J. and Edward J. Vytlacil**, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 2005, *73* (3), 669–738.

**Heckman, James J, Margaret L Holland, Kevin K Makino, Rodrigo Pinto, and Maria Rosales-Rueda**, "An Analysis of the Memphis Nurse-Family Partnership Program," Working Paper 23610, National Bureau of Economic Research 2017.

**Imai, Kosuke, Luke Keele, and Teppei Yamamoto**, "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 2010, *25* (1), 51–71.

**Imbens, Guido W. and Charles F. Manski**, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72* (6), 1845–1857.

_ **and Joshua D. Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

**Kalil, Ariel, Rebecca Ryan, and Michael Corey**, "Diverging destinies: Maternal education and the developmental gradient in time with children," *Demography*, 2012, *49* (4), 1361–1383.

**Kline, Patrick and Christopher R. Walters**, "On Heckits, LATE, and Numerical Equivalence," *Econometrica*, 2019, *87* (2), 677–696.

**Kowalski, Amanda E.**, "How to Examine External Validity Within an Experiment," *Journal of Economics & Management Strategy*, 2023, *32* (3), 491–509.

**Love, John M, Ellen Eliason Kisker, Christine Ross, Helen Raikes, Jill Constantine, Kimberly Boller, Jeanne Brooks-Gunn, Rachel Chazan-Cohen, Louisa Banks Tarullo, Christy Brady-Smith et al.**, "The effectiveness of Early Head Start for 3-year-old children and their parents: lessons for policy and programs," *Developmental Psychology*, 2005, *41* (6), 885–901.

**Luxen, Dennis and Christian Vetter**, "Real-Time Routing with OpenStreetMap Data," in "Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems" Association for Computing Machinery 2011, pp. 513–516.

**Manski, Charles F.**, "Monotone Treatment Response," *Econometrica*, 1997, *65* (6), 1311–1334.

_ , *Partial Identification of Probability Distributions* Springer Series in Statistics, New York: Springer-Verlag, 2003.

_ **and John V. Pepper**, "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 2000, *68* (4), 997–1010.

**Meghir, Costas, Orazio Attanasio, Pamela Jervis, Monimalika Day, Prerna Makkar, Jere Behrman, Prachi Gupta, Rashim Pal, Angus Phimister, Nisha Vernekar, and Sally Grantham-McGregor**, "Early Stimulation and Enhanced Preschool: A Randomized Trial," *Pediatrics*, May 2023, *151* (Supplement 2), e2023060221H.

**Mogstad, Magne, Andres Santos, and Alexander Torgovitsky**, "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 2018, *86* (5), 1589–1619.

**Putnam, Robert D**, *Our kids: The American dream in crisis*, Simon and Schuster, 2016.

**Reardon, Sean F and Ximena A Portilla**, "Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry," *AERA Open*, 2016, *2* (3), 2332858416657343.

**Ryder, Harl E. Jr. and Geoffrey M. Heal**, "Optimal Growth with Intertemporally Dependent Preferences," *Review of Economic Studies*, 1973, *40* (1), 1–31.

**Skrondal, Anders and Petter Laake**, "Regression among factor scores," *Psychometrika*, 2001, *66* (4), 563–575.

**St. Pierre, Robert G., Anne Ricciuti, Fumiyo Tao, Cindy Creps, Janet Swartz, Wang Lee, Amanda Parsad, and Tracy Rimdzius**, "Third National Even Start Evaluation: Program Impacts and Implications for Improvement," Technical Report, U.S. Department of Education, Planning and Evaluation Service, Elementary and Secondary Division 2003.

**Todd, Petra E. and Kenneth I. Wolpin**, "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 2006, *96* (5), 1384–1417.

**Todd, Petra E and Kenneth I Wolpin**, "The production of cognitive achievement in children: Home, school, and racial test score gaps," *Journal of Human Capital*, 2007, *1* (1), 91–136.

**U.S. Census Bureau**, "Census Geocoder Documentation," https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation/census-geocoder.html 2026. Accessed August 8, 2026.

**Vytlacil, Edward**, "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 2002, *70* (1), 331–341.

**Wagner, Mary M. and Serena L. Clayton**, "The Parents as Teachers program: Results from two demonstrations," *The Future of Children*, 1999, *9* (1), 91–115.

**Westheimer, Miriam**, *Parents Making a Difference: International Research on the Home Instruction for Parents of Preschool Youngsters (HIPPY) Program*, The Hebrew University Magnes Press, 2003.

**Zhou, Jin, James J. Heckman, Bei Liu, Mai Lu, Susan M. Chang, and Sally Grantham-McGregor**, "Comparing China REACH and the Jamaica Home Visiting Program," *Pediatrics*, May 2023, *151* (Supplement 2), e2023060221I.

# Appendix A The JumpStart Program

## A.1 Curriculum

Table A1: JumpStart Program Curriculum

| Session | Month | Child Present? | Book | CD | Other Materials |
|---|---|---|---|---|---|
| 1 | October | No | No, David! | The Color Songs | Play Dough, Laminated Sentence Strip, Laminated Construction Mat, Expo Marker |
| 2 | October | No | The Pigeon Needs a Bath | The Color Songs, Brush a Brush | Play Dough, Laminated Sentence Strip, Laminated Construction Mat, Expo Marker, Hand Pump Soap, Tooth Brush, Tooth Paste, Small Pack of Tissues |
| 3 | October | Yes | No, David! | Name Song, Brush a Brush | Hand Sanitizer, Pre-made Laminated Sentence Strip with Child's Name |
| 4 | November | No | Brown Bear, Brown Bear | The Color Songs | Red Construction Paper, Glue Bottles, Leaf Outline, Apple Outline |
| 5 | November | No | Brown Bear, Brown Bear | The Color Songs | Blue Construction Paper, Glue Bottles, Butterfly Outline, Bird Outline |
| 6 | November | Yes | Brown Bear, Brown Bear | The Color Songs | Pre-made Laminated Sentence Strip with Child's Name Written in Blue and Red, Red and Blue Beads, Red and Blue Markers |
| 7 | December | No | Chicka Chicka, Boom Boom | The Color Songs | Scissors, Green, Red, and Blue Strips of Construction Paper |
| 8 | December | No | Dr. Seuss's ABC: An Amazing Alphabet Book! | The Color Songs | Play Dough, Scissors, Green Construction Paper, Green Sticker Dots, Frog Outline, Alligator Outline, Glue |
| 9 | December | Yes | Dr. Seuss's ABC: An Amazing Alphabet Book! | The Color Songs | Pre-made Laminated Sentence Strip with Child's Name Written in Green, Tree Outline, Scissors, Green, Blue, and Red Sticker Dots |
| 10 | January | No | Rainbow Fish | The Color Songs, Five Little Fish, Five Little Monkeys | Large Crayons, White Copy Paper, Three White Die-Cut Fish, Yellow Construction Paper, Ocean Scene Outline, Cotton Balls, Glue |
| 11 | January | No | Ten Black Dots | The Color Songs, Five Little Fish, Five Little Monkeys | Large Crayons, Dot-to-Dot Paper, Spider-to-Dot Paper, Cutting Line Paper, Spider Web, Five Black Spiders |
| 12 | January | Yes | Ten Black Dots | The Color Songs, Five Little Monkeys | Pre-made Laminated Sentence Strip with Child's Name Written in Black, Bee Outline, Yellow Construction Paper, Scissors, Glue, Wax Paper Wings, Black Crayon, Black Sticker Dots |
| 13 | February | No | Harold and the Purple Crayon | The Color Songs | Tongs, Pom Poms, Sorting Mat, Clear Plastic Cup |
| 14 | February | No | The Three Bears | The Color Songs | Goldilocks Cutting Paper, Play Dough, Die Cut Counting Bears, Sorting Mat |
| 15 | February | Yes | The Three Bears | The Color Songs | Pre-made Laminated Sentence Strip with Child's Name Written in Purple, Goldilocks Rhyme, Popsicle Sticks, Bear Outline, Goldilocks Die Cut, Purple Marker |
| 16 | March | No | From Head to Toe | The Color Songs, Head, Shoulders, Knees, and Toes | Simon Says Poster |
| 17 | March | No | If You're Happy and You Know It | The Color Songs, Head, Shoulders, Knees, and Toes | Circle Outline, Glue, Scissors, Crayons, Black and Blue Dots, Copy Paper, Play Dough, Counting Mat, Brown Construction Paper |
| 18 | March | Yes | From Head to Toe | The Color Songs, Head, Shoulders, Knees, and Toes | Pre-made Laminated Sentence Strip with Child's Name Written in Orange |
| 19 | April | No | Three Little Pigs | The Color Songs | Beads, Yarn, Cut-Up Straws, Pipe Cleaners, Buttons, Pink Construction Paper, Crayons, Pig Outline |
| 20 | April | No | Three Little Pigs | The Color Songs, Five Little Monkeys | Spinning Tops, Play Coins, Crayons, Pink Coloring Page, Cookie Cutters, Sentence Strip |
| 21 | April | Yes | Three Little Pigs | The Color Songs, Head, Shoulders, Knees, and Toes | Pre-made Laminated Sentence Strip with Child's Name Written in Pink, Name Puzzle, Pink Construction Paper, Glue |
| 22 | May | Yes | Mary Had a Little Lamb Rhyme Sheet, Humpty Dumpty Rhyme Sheet | The Color Songs | Lamb Outline, Glue, Cotton Balls, Tweezers, Counting Mat, Red Construction Paper, Crayons, Scissors, Egg Outline, Black Crayon, Humpty Dumpty's Wall Outline |

## A.2 Program Cost

The cost of JumpStart combines three inputs, family liaison time, parental time, and materials. Each liaison spends two hours on each of the 22 sessions, one preparing and one leading the meeting with parents, or 44 hours over the program for a group of about ten families. At an hourly wage of $9, liaison time costs about $40 per family. Our data indicate that parents spend an additional 6.6 hours interacting with their children, and they spend one hour attending each session, or 22 hours in total, for a time commitment of 28.6 hours. Log wage regressions that control for selection into employment put the parents' opportunity cost at $7.25 per hour, so parental time costs about $207 per family. Travel to the sessions adds little and is left out of this figure, since the median family in the lottery lives a 2.5-minute drive from its school (mean 2.9 minutes, 99th percentile 11.6), so that a round trip by car adds about six minutes per session on average and about $16 per family over the program. Materials cost $200 per family. The total is therefore $447, a little under $450 per family per year.

# Appendix B Additional Details of the Experimental Design

## B.1 Program History, Recruitment, and Sample Detail

This subsection provides the recruitment detail behind Section 3.1 and Table 1.

Before the evaluation, about eight district elementary schools offered JumpStart to families in their catchment areas, with total enrollment of roughly 60 families across the district. Each school covered the materials cost of $200 per family from its own budget, and this cost was a major impediment to schools adopting or expanding the program. As part of the evaluation, the Laura and John Arnold Foundation funded the expansion of the program to all 24 elementary schools, 22 of which took part in the third year. Before the study, liaisons recruited families first-come, first-served, ending when the school's budget-determined target was reached; the lottery replaced this allocation for the evaluation.

The year-one strategy assumed sufficient interest in all 24 catchment areas to meet a target of 20 lottery families per school, 10 to be assigned to each arm, supported by an individual liaison incentive for reaching the target. Twenty of the 24 schools met it, and 461 families entered the lottery. Enrollment success, however, masked uneven interest: in 15 of the 24 schools at least one family assigned a program spot never attended a session; overall, 12 percent of treated lottery families never attended and 27 percent attended five or fewer of the 22 sessions. In years two and three the incentive was restructured, an individual incentive at 14 families with a group incentive for additional recruits. In year two, 389 families entered the lottery across 24 schools; only 4 schools had a family that never attended, 2 percent of treated families never attended, and 16 percent attended five or fewer sessions. In year three, two liaison departures just before the school year reduced coverage to 22 schools, and 366 families entered the lottery; attendance weakened for unclear reasons, with 8 percent never attending and 22 percent attending five or fewer sessions, worse than year two though not as severe as year one.

Table B1 reports the full sample accounting. Study consent at baseline was high in all years: 90.9 percent of treatment and 89.5 percent of control lottery families in year one, 92.5 and 93.7 percent in year two, and 94.2 and 96.0 percent in year three; cumulatively, 1,126 of 1,216 lottery families consented, a refusal rate of 7.4 percent. At endline, about eight months after the start of the program, re-interview rates among baseline consenters rose across cohorts, from 71.4 percent in year one to 82.3 and 84.8 percent in years two and three. Attrition was higher in the control group, and the differential fell over time: computed on the base of baseline consenters, the control-group attrition rate exceeded the treatment-group rate by 18.6 percentage points in year one, 9.4 in year two, and 6.1 in year three. Section 4.5 and Appendix C.6 assess the consequences.

## B.2 Randomization Balance

Table B3 compares treatment and control families on sixteen baseline characteristics, in the full sample of families consenting at baseline, the experiment sample, and the model sample, includ-

Table B1: Sample Sizes

| | Cohort 1 | | Cohort 2 | | Cohort 3 | | All | |
|---|---|---|---|---|---|---|---|---|
| | Treatment | Control | Treatment | Control | Treatment | Control | Treatment | Control |
| **In lottery** | | | | | | | | |
| Observations | 232 | 229 | 199 | 190 | 190 | 176 | 621 | 595 |
| % in treatment/control group | 50.33 | 49.67 | 51.16 | 48.84 | 51.91 | 48.09 | 51.07 | 48.93 |
| **In study** | | | | | | | | |
| *Baseline* | | | | | | | | |
| Observations | 211 | 205 | 184 | 178 | 179 | 169 | 574 | 552 |
| % in treatment/control group | 50.72 | 49.28 | 50.83 | 49.17 | 51.44 | 48.56 | 50.98 | 49.02 |
| % of lottery | 90.95 | 89.52 | 92.46 | 93.68 | 94.21 | 96.02 | 92.43 | 92.77 |
| *Endline (experiment sample)* | | | | | | | | |
| Observations | 170 | 127 | 160 | 138 | 157 | 138 | 487 | 403 |
| % in treatment/control group | 57.24 | 42.76 | 53.69 | 46.31 | 53.22 | 46.78 | 54.72 | 45.28 |
| % of lottery | 73.28 | 55.46 | 80.40 | 72.63 | 82.63 | 78.41 | 78.42 | 67.73 |
| % of baseline in-study | 80.57 | 61.95 | 86.96 | 77.53 | 87.71 | 81.66 | 84.84 | 73.01 |
| *Model sample* | | | | | | | | |
| Observations | 155 | 119 | 153 | 135 | 146 | 134 | 454 | 388 |
| % in treatment/control group | 56.57 | 43.43 | 53.12 | 46.88 | 52.14 | 47.86 | 53.92 | 46.08 |
| % of lottery | 66.81 | 51.97 | 76.88 | 71.05 | 76.84 | 76.14 | 73.11 | 65.21 |
| % of endline in-study | 91.18 | 93.70 | 95.62 | 97.83 | 92.99 | 97.10 | 93.22 | 96.28 |

*Note:* Families by cohort and treatment status. *In lottery* is all families randomized; *In study* is those who consented at baseline and, at endline, those re-interviewed about eight months after the start of the program; *Model sample* is the subset of the endline sample used to estimate the structural model of Section 5. Shares are of the treatment-control total within each row block, of the corresponding lottery sample (% of lottery), of the baseline in-study sample for the endline rows (% of baseline in-study), and of the endline in-study sample for the model rows (% of endline in-study). All families entering the lottery were randomized regardless of study consent.

ing the two blocking variables (baseline JumpStart score and child age). No difference is significant at the 5 percent level in any of the three samples. The smallest $p$-value, 0.09, belongs to home-to-school distance, which was not a blocking variable; with sixteen covariates, one marginal $p$-value is within sampling expectation, and the estimation of Section 5 uses distance only within the offered arm, where the graduation decision is modeled, never as a treatment–control contrast.

## B.3 Measurement

This subsection details the measures summarized in Section 3.2: the two test instruments, the survey measures of parental involvement, and the three constructed measures used in the analysis (the economic well-being index, the parental-investment factor scores, and the home-to-campus driving distance).

**JumpStart test score.** The AISD JumpStart Assessment Tool takes about 10–15 minutes and measures the skills the JumpStart curriculum teaches, with 26 questions in six sub-modules, colors (10 questions), fine motor skills (8), counting (2), sorting (1), name recognition (1), and book handling (4).

Table B2: Summary Statistics

| | Baseline | | | Endline | | |
|---|---|---|---|---|---|---|
| | *N* | Mean | SD | *N* | Mean | SD |
| **JumpStart test outcomes** | | | | | | |
| All modules (26) | | | | | | |
| *Test score (% correct)* | 890 | 50.96 | 25.72 | 890 | 72.75 | 20.90 |
| *% lower than 25%* | 890 | 19.66 | 39.77 | 890 | 2.92 | 16.85 |
| *% higher than 75%* | 890 | 20.34 | 40.27 | 890 | 52.81 | 49.95 |
| Module-specific % score (no. of items): | | | | | | |
| *Colors (10)* | 890 | 58.91 | 37.99 | 890 | 76.71 | 28.71 |
| *Fine motor skills (8)* | 890 | 56.54 | 29.51 | 890 | 77.75 | 21.56 |
| *Counting (2)* | 890 | 42.02 | 43.84 | 890 | 64.72 | 42.32 |
| *Sorting (1)* | 890 | 43.26 | 49.57 | 890 | 70.79 | 45.50 |
| *Name recognition (1)* | 890 | 24.83 | 43.23 | 890 | 59.66 | 49.08 |
| *Book handling (4)* | 890 | 32.81 | 31.87 | 890 | 60.62 | 38.27 |
| **Bracken test outcomes** | | | | | | |
| All modules (no. of items = 85) | | | | | | |
| *Test score (% correct)* | 890 | 23.86 | 19.02 | 888 | 35.66 | 18.74 |
| *Percentile score* | 889 | 27.09 | 28.99 | 888 | 28.40 | 26.48 |
| *% delayed or very delayed* | 889 | 50.73 | 50.02 | 888 | 44.59 | 49.73 |
| Module-specific % score (no. of items): | | | | | | |
| *Colors (10)* | 890 | 57.75 | 41.62 | 890 | 80.28 | 31.77 |
| *Letters (15)* | 890 | 21.27 | 28.98 | 890 | 30.75 | 30.96 |
| *Numbers (18)* | 890 | 13.23 | 23.14 | 890 | 22.10 | 27.60 |
| *Shapes (20)* | 890 | 22.53 | 21.21 | 888 | 36.06 | 22.59 |
| *Sizes (22)* | 890 | 20.12 | 20.04 | 889 | 29.65 | 18.81 |
| **Parental inputs outcomes** | | | | | | |
| No. days/week read to the child | 885 | 3.26 | 2.54 | 887 | 4.46 | 2.16 |
| No. hours/week read to the child | 882 | 1.19 | 1.53 | 886 | 1.59 | 1.40 |
| No. of children's books at home | 886 | 6.43 | 3.49 | 887 | 8.10 | 2.69 |
| Helps child learn: | | | | | | |
| *Alphabet* | 888 | 82.09 | 38.36 | 887 | 89.18 | 31.08 |
| *Colors* | 885 | 89.04 | 31.26 | 887 | 97.52 | 15.56 |
| *Numbers* | 887 | 87.26 | 33.36 | 887 | 95.72 | 20.26 |
| *Shapes* | 882 | 67.46 | 46.88 | 887 | 81.29 | 39.03 |
| No. of skills helped with at home (0–4) | 880 | 3.26 | 1.12 | 887 | 3.64 | 0.68 |
| No. of hours/day child exposed to TV | 872 | 3.15 | 2.61 | 881 | 3.49 | 2.83 |
| No. days/month takes child to grocery store | 886 | 5.61 | 2.58 | 887 | 5.79 | 2.40 |
| Parental inputs factor score | 862 | 1.20 | 0.40 | 880 | 1.58 | 0.98 |
| **Demographics** | | | | | | |
| Economic well-being index (0–9) | 858 | 7.62 | 1.53 | | | |
| Parent years of schooling | 881 | 12.24 | 3.52 | | | |
| Parent Hispanic (%) | 888 | 60.81 | 48.84 | | | |
| Parent age (years) | 803 | 32.60 | 6.55 | | | |
| No. of children in the household | 886 | 2.85 | 1.17 | | | |
| Child male (%) | 890 | 49.44 | 50.02 | | | |
| Child age (years) | 890 | 3.52 | 0.30 | | | |
| Home-school distance (km) | 880 | 1.83 | 2.54 | | | |

*Note:* Number of observations, mean and standard deviation of each variable at baseline and at endline, in the experiment sample. Test scores are in percent of items answered correctly, with the number of items in parentheses after each module. The percentile score and the share classified as delayed or very delayed are measured against the Bracken normalizing national sample. The parental inputs factor score is in weekly reading hours (Appendix B.4). Demographics are recorded at baseline only; the economic well-being index runs from 0 to 9 (Appendix B.5).

Table B3: Treatment-Control Balance

| | Full Sample | | | Experiment Sample | | | Model Sample | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control mean | T-C diff. | *p*-value | Control mean | T-C diff. | *p*-value | Control mean | T-C diff. | *p*-value |
| **Blocking variables** | | | | | | | | | |
| JumpStart baseline score (%) | 50.95 | 0.43 | 0.78 | 50.45 | 0.93 | 0.59 | 50.39 | 0.20 | 0.91 |
| Child age (years) | 3.52 | 0.01 | 0.65 | 3.52 | –0.00 | 0.96 | 3.52 | 0.00 | 0.97 |
| **Non-blocking variables** | | | | | | | | | |
| Bracken baseline score (%) | 23.93 | 1.05 | 0.36 | 23.05 | 1.48 | 0.25 | 23.00 | 1.37 | 0.29 |
| No. hours/week read to the child | 1.19 | 0.04 | 0.68 | 1.14 | 0.10 | 0.35 | 1.12 | 0.09 | 0.35 |
| No. of children's books at home | 6.46 | –0.07 | 0.75 | 6.48 | –0.10 | 0.67 | 6.48 | –0.07 | 0.78 |
| No. of skills helped with at home (0–4) | 3.27 | –0.03 | 0.68 | 3.26 | –0.00 | 0.98 | 3.27 | –0.02 | 0.83 |
| No. of hours/day child exposed to TV | 3.21 | –0.08 | 0.60 | 3.24 | –0.15 | 0.40 | 3.25 | –0.14 | 0.43 |
| No. days/month takes child to grocery store | 5.64 | –0.06 | 0.69 | 5.66 | –0.10 | 0.56 | 5.67 | –0.13 | 0.46 |
| Parental inputs factor score | 1.21 | –0.01 | 0.69 | 1.20 | –0.00 | 0.96 | 1.20 | –0.00 | 0.96 |
| Economic well-being index (0–9) | 7.59 | –0.02 | 0.84 | 7.63 | –0.02 | 0.85 | 7.63 | –0.02 | 0.87 |
| Parent years of schooling | 12.05 | 0.21 | 0.33 | 12.06 | 0.32 | 0.18 | 12.08 | 0.31 | 0.20 |
| Parent Hispanic (%) | 59.71 | –2.54 | 0.39 | 62.28 | –2.70 | 0.41 | 62.63 | –2.23 | 0.51 |
| Parent age (years) | 32.16 | 0.40 | 0.34 | 32.50 | 0.19 | 0.69 | 32.50 | 0.18 | 0.71 |
| No. of children in the household | 2.79 | 0.11 | 0.12 | 2.81 | 0.08 | 0.31 | 2.80 | 0.09 | 0.26 |
| Child male (%) | 49.28 | –0.67 | 0.82 | 47.64 | 3.28 | 0.33 | 47.94 | 3.16 | 0.36 |
| Home-school distance (km) | 1.74 | 0.25 | 0.09 | 1.70 | 0.25 | 0.14 | 1.70 | 0.22 | 0.21 |

*Note:* Control-group mean, treatment-control difference (T-C diff.) and the *p*-value of the difference, in each of the three samples the paper uses, the full sample of families consenting at baseline, the experiment sample of families observed at both baseline and endline, and the sub-sample used to estimate the structural model. The JumpStart baseline score and child age are the two blocking variables; all remaining variables were not used in the randomization. Units in parentheses.

**Bracken test score.** The evaluation also used the Bracken School Readiness Assessment, Third Edition (BSRA-3) (Bracken, 2007), an individual cognitive test for children from pre-K through second grade that measures a child's exposure to concepts necessary for learning at school. BSRA-3 includes five subtests (85 questions in total), all reflecting receptive and expressive language skills: colors (10 questions, identifying common colors by name); letters (15, identifying upper- and lower-case letters); numbers and counting (18, identifying single- and double-digit numerals and counting objects); sizes and comparisons (22, understanding words such as tall or wide and matching or differentiating objects on a characteristic); and shapes (20, identifying basic shapes by name). The test is administered individually, with oral and visual prompts, so its length depends on the child's performance; under the discontinuation rule, after five consecutive incorrect responses within a module the remaining questions in that module are scored as incorrect and the assessment proceeds to the next module. At both baseline and endline, each correct response receives one point, summed across modules to a raw score from 0 to 85. From the raw score we construct three measures: the proportion of correct responses (raw score divided by 85); an age-adjusted percentile score mapped from the raw score and the child's age category; and an indicator for children classified as delayed or very delayed under the criteria of Bracken (2007) (for example, among children aged 36–38 months, a raw score of 12 or lower is classified as delayed or very delayed).

**Parental involvement.** Families in the study completed a baseline and an endline survey with questions on the extent to which family members engaged in learning activities with the child:

the number of days per week and total hours per week spent reading to the child, the number of children's books in the household, and whether family members helped the child learn the alphabet, colors, numbers, and shapes. The surveys also recorded the number of hours per day the television was on while the child was at home and how frequently the parent took the child to the grocery store.

### B.4 Construction of the Investment Measure

Each of the five parental investment items captures a distinct dimension of the home learning environment and is measured with error. We therefore summarize them by the score on a single common factor, estimated separately at baseline and at endline so that the two waves are not constrained to share a loading structure. This subsection sets out the three steps of that construction: the factor model, its rotation, and the scoring and anchoring of the predicted factor.

**Exploratory factor analysis.** We estimate an exploratory factor analysis (EFA) of (i) the number of hours per week the parent reads to the child, the product of days per week and minutes per occasion, top-coded at 10.5 hours; (ii) the number of school-readiness skills someone at home helps the child with, from 0 to 4 (the alphabet, colors, numbers and shapes); (iii) the number of children's books in the home, coded at the midpoint of the reported category (0, 1.5, 6 or 10); (iv) the number of hours per day the television is on while the child is at home, top-coded at 12; and (v) the number of grocery trips per month on which the child is taken, coded from the reported frequency (8, 4, 1 or 0). The estimation retains two factors in each wave, only the first of which loads appreciably on the reading, skills and books items.

We then rotate the factors obliquely, by promax. An oblique rotation drives each item toward a single factor, as an orthogonal rotation does, but without forcing the factors themselves to be uncorrelated, a restriction that would be implausible here, since the dimensions of parental engagement are related in practice. Table B4 reports the loadings before and after the rotation, together with the uniquenesses, which the rotation leaves unchanged. The rotated first factor loads most heavily on the weekly hours the parent reads to the child, the number of skills helped with at home and the number of children's books, in both waves; television exposure loads close to zero, grocery trips only weakly, at 0.26 at baseline and 0.07 at endline, and both have uniquenesses above 0.92, so they contribute almost nothing to the score.

**Factor scores and anchoring.** Predicted scores on the first factor are then formed by the regression formula at baseline and by the Bartlett formula at endline, following Skrondal and Laake (2001). The location and scale of a latent factor are arbitrary, and we fix them by anchoring the factor on weekly reading hours. The factor model is estimated on standardized items, so that item $j$ in wave $t$ satisfies

$$M_{ijt} = \mu_{jt} + s_{jt}\,\lambda_{jt}\,F_{it} + s_{jt}\,e_{ijt},$$

Table B4: Factor Loadings Before and After Rotation, and Uniquenesses

| | Before Rotation | | After Rotation | | |
|---|---|---|---|---|---|
| Measure | Factor 1 | Factor 2 | Factor 1 | Factor 2 | Uniqueness |
| **Panel A: Baseline** | | | | | |
| No. of hours/week reading to child | 0.406 | –0.033 | 0.413 | –0.033 | 0.834 |
| No. of skills helped with at home (0–4) | 0.486 | 0.001 | 0.486 | 0.001 | 0.763 |
| No. of children's books at home | 0.465 | 0.006 | 0.464 | 0.006 | 0.784 |
| No. of hours/day child exposed to TV | –0.010 | 0.117 | –0.036 | 0.120 | 0.986 |
| No. of grocery trips/month with child | 0.268 | 0.043 | 0.258 | 0.044 | 0.926 |
| **Panel B: Endline** | | | | | |
| No. of hours/week reading to child | 0.391 | 0.010 | 0.388 | –0.014 | 0.847 |
| No. of skills helped with at home (0–4) | 0.341 | 0.074 | 0.355 | 0.054 | 0.878 |
| No. of children's books at home | 0.355 | –0.023 | 0.344 | –0.045 | 0.874 |
| No. of hours/day child exposed to TV | –0.151 | 0.170 | –0.106 | 0.182 | 0.948 |
| No. of grocery trips/month with child | 0.024 | 0.191 | 0.071 | 0.193 | 0.963 |

*Note:* Exploratory factor analysis of the five parental-investment measures of Section 3.2, estimated separately at baseline (Panel A) and endline (Panel B). Columns (1) and (2) report the unrotated loadings on the first two factors and columns (3) and (4) the loadings after rotation. Column (5) reports the uniqueness of each measure, one minus its communality, the sum of its squared unrotated loadings in columns (1) and (2), which the rotation leaves unchanged. The investment indices $X_{i0}$ and $X_{i1}$ used in Section 5 are the scores on the first factor, constructed as described below and measured in hours of weekly reading.

where $\mu_{jt}$ and $s_{jt}$ are the sample mean and standard deviation of the item, $\lambda_{jt}$ is its rotated loading on the first factor, and $F_{it}$ has mean zero and unit variance. Defining the investment index as $X_{it} \equiv \mu_{at} + s_{at}\lambda_{at}F_{it}$, with $a$ denoting weekly reading hours, the anchoring item becomes $M_{iat} = X_{it} + s_{at}e_{iat}$, with intercept zero and loading one. We apply the same linear map to the predicted factor, $\widehat{X}_{it} = \mu_{at} + s_{at}\lambda_{at}\widehat{F}_{it}$, for both the regression and the Bartlett score, so that each keeps its properties. One point of the index is the change in investment that raises expected weekly reading time by one hour. Anchoring both waves on the same item puts the two indices in a common unit, expected weekly reading time, under the assumption that reading time relates to latent investment in the same way at baseline and at endline.[40]

**Why two scoring methods.** Neither score equals the latent index, and the two err in opposite ways. The Bartlett score is conditionally unbiased, $\widehat{X}^B_{it} = X_{it} + \epsilon^B_{it}$ with $\mathbb{E}[\epsilon^B_{it} \mid X_{it}] = 0$, so its error behaves like classical measurement error. The regression score is the best linear predictor of the index given the items, $X_{it} = \widehat{X}^R_{it} + \epsilon^R_{it}$ with $\epsilon^R_{it}$ uncorrelated with the items and hence with the score itself. With a single factor, and in deviations from the mean, it is the Bartlett score shrunk by the reliability of the items,

$$\widehat{X}^R_{it} = r_t\,\widehat{X}^B_{it}, \qquad r_t \equiv \frac{\mathrm{Var}(X_{it})}{\mathrm{Var}(X_{it}) + \mathrm{Var}(\epsilon^B_{it})} < 1.$$

[40] These are the units in which $\gamma_G$ and $\beta_{X_1}$ are reported in Table 4.

Skrondal and Laake (2001) show that these properties dictate where each score belongs. Suppose first that the index enters as a regressor, $Y_i = \beta X_{it} + u_i$. Substituting each score gives

$$Y_i = \beta \widehat{X}^R_{it} + \big(u_i + \beta \epsilon^R_{it}\big) \qquad \text{and} \qquad Y_i = \beta \widehat{X}^B_{it} + \big(u_i - \beta \epsilon^B_{it}\big).$$

In the first equation the composite error is uncorrelated with $\widehat{X}^R_{it}$, so least squares recovers $\beta$. In the second it is correlated with $\widehat{X}^B_{it}$, which contains $\epsilon^B_{it}$, and least squares converges to $r_t\beta$, the familiar attenuation. Suppose instead that the index is the dependent variable, $X_{it} = \gamma Z_i + v_i$. Then

$$\widehat{X}^B_{it} = \gamma Z_i + \big(v_i + \epsilon^B_{it}\big) \qquad \text{and} \qquad \widehat{X}^R_{it} = r_t \gamma Z_i + r_t\big(v_i + \epsilon^B_{it}\big).$$

The Bartlett error only adds to the disturbance, so least squares recovers $\gamma$, whereas the regression score scales the coefficient down to $r_t\gamma$. The single-factor case only makes the argument transparent, and Skrondal and Laake (2001) show that it goes through unchanged with several correlated factors, as in our model. In the investment equation $X_{i1}$ is the dependent variable and $X_{i0}$ a regressor, hence the Bartlett score at endline and the regression score at baseline. With other regressors in the equation, the result for the regression score also requires $\epsilon^R_{i0}$ to be uncorrelated with them. Randomization guarantees this for the offer, but it holds only approximately for graduation and the controls, which do not enter the factor model.

## B.5 Construction of the Economic Well-being Index

The surveys contain no reliable continuous measure of household resources, since income is reported in bracketed categories and is missing or refused for a substantial share of families. We therefore summarize a family's economic security by an index built from eight categorical items that were asked identically in both waves and that record material hardship, reliance on transfers, and food insecurity. The index is constructed the same way at baseline and at endline, so that changes over the program year are measured on a common scale.

The index has three components, each oriented so that higher values mean greater economic security. The first records the utilities that have *not* been disconnected. Parents report whether, in the three months before the interview, the household's telephone, electricity, gas or water was ever cut off, and we subtract the number cut off from four, so the component runs from 0, when all four had been cut off, to 4, when none had. The second records reliance on transfers. Parents report whether the household receives food stamps, public assistance or welfare, and child support, and we subtract the number of sources received from three, so the component runs from 0 for a family drawing on all three to 3 for a family drawing on none. The third records food insecurity, from the question of how often the parent worries that food will run out before there is money to buy more, whose three answers, often true, sometimes true, and never true, are scored 0, 1 and 2 respectively.

The overall index is the sum of the three components,

$$\text{Economic well-being}_i = \underbrace{\big(4 - \#\{\text{utilities cut off}\}\big)}_{0 \text{ to } 4} + \underbrace{\big(3 - \#\{\text{transfer sources}\}\big)}_{0 \text{ to } 3} + \underbrace{\text{food-worry score}}_{0 \text{ to } 2},$$

so that it runs from 0 to 9 by construction. A family is assigned a value only when all eight underlying items are answered; families missing any item are treated as missing on the index and are handled by the mean-imputation-plus-indicator convention applied to the other controls. In our data the index takes values from 1 to 9 and averages 7.63 in the control group of the experiment sample at baseline (Table B3); it is balanced across the treatment and control arms in every cohort and in both the full and the experiment sample.

## B.6 Construction of the Home-to-Campus Driving Distance Measure

The distance measure is the driving distance, in kilometers, from a family's home to the campus at which its JumpStart sessions were held. It is built in two steps: the addresses are converted to coordinates, and the coordinates are routed over a road network contemporaneous with the family's cohort.

We first clean and standardize the residential addresses reported on the program applications and the addresses of the AISD campuses, and geocode them with the U.S. Census Batch Geocoder (U.S. Census Bureau, 2026), which returns the latitude and longitude that serve as origins and destinations for the routing step. For the small number of addresses the Census geocoder cannot match, we substitute fallback coordinates, taken either from a manual Google Maps review or

from the Census representative point for the ZIP area, which are frozen and documented so that the measure is reproducible.

We then compute driving distances with the Open Source Routing Machine (OSRM) (Luxen and Vetter, 2011), an open-source engine that routes over OpenStreetMap road-network data.[41] Unlike straight-line distance, OSRM follows the drivable network and respects road access and turn restrictions. To match the network to the timing of the intervention, we route each cohort over an archived Texas map snapshot dated January 1 of 2016, 2017 and 2018 for cohorts 1, 2 and 3 respectively. The primary measure is the recommended car route produced by OSRM's standard car profile, which favors the roads and maneuvers of ordinary automobile travel; as an alternative we also compute the shortest routable car distance over the same network.

Among the 1,205 of the 1,216 lottery families for which a distance could be constructed, the recommended-route distance averages 1.87 kilometers, with a standard deviation of 2.42 kilometers and a median of 1.40 kilometers, and the interquartile range is 0.86 to 2.12 kilometers. Families live close to their campus, as 71.9 percent are within 2 kilometers, 95.8 percent within 5 kilometers and 99.0 percent within 10 kilometers. The alternative shortest-route measure averages 1.76 kilometers and has a correlation of 0.996 with the primary measure, so nothing in the analysis turns on the choice between them.

### B.7 Program Attendance

This subsection details the JumpStart attendance patterns. Each month of the program has three group sessions at the family's local elementary school: two with parents alone and one attended by parents and children together. Missed sessions were eligible for make-up, and families were dropped from the program after three consecutive unexcused absences. Table B5 reports the distribution of completed sessions, inclusive of make-ups, for treated families in the experiment sample. Mean completed sessions were 16.3, 19.0, and 17.7 across the three cohorts, 17.6 pooled. Graduation, defined as completing 16 or more of the 22 sessions, was 70.6 percent in the first cohort, in large part a consequence of the initial recruitment strategy (Appendix B.1), and 84.4 and 80.3 percent after its revision, 78.2 percent pooled. Non-completion concentrates at the very bottom of the distribution, though unevenly across cohorts: among non-graduates, 60 percent completed 5 sessions or fewer in the first cohort and 65 percent in the third, against 44 percent in the second. Because the experiment sample conditions on completing the endline assessment, these figures are somewhat more favorable than the lottery-sample attendance of Appendix B.1: the families that attended least were also the most likely to attrit before endline.

[41] OSRM documentation is available at https://project-osrm.org/docs/.

Table B5: JumpStart Program Completion by Cohort

| | Cohort 1 | Cohort 2 | Cohort 3 | Pooled |
|---|---|---|---|---|
| **Mean completed sessions** | 16.25 | 18.96 | 17.66 | 17.59 |
| **Completion rates (%)** | | | | |
| 0 sessions | 5.88 | 0.62 | 6.37 | 4.31 |
| 1–5 sessions | 11.76 | 6.25 | 6.37 | 8.21 |
| 6–15 sessions | 11.76 | 8.75 | 7.01 | 9.24 |
| 16+ sessions | 70.59 | 84.38 | 80.25 | 78.23 |

*Note:* Sessions of the 22-session curriculum completed by treated families in the experiment sample, inclusive of made-up classes, by cohort. *Mean completed sessions* is the average number of curriculum sessions completed; *Completion rates* are the shares attending 0, 1–5, 6–15 and 16 or more sessions. Graduation from the program requires 16 or more sessions. Missed sessions were eligible for make-up; families were dropped from the program after three consecutive unexcused absences.

## B.8 Characteristics of Graduates and Non-Graduates

Table B6 compares treated families that completed the JumpStart program (at least 16 of the 22 sessions) with those that did not, on baseline characteristics in the experiment sample; the main text summarizes the pattern in Section 3.4. Graduates enter with significantly higher baseline test scores: 53.2 versus 44.9 percent on the JumpStart test (a difference of 8.3 percentage points, $p < 0.01$) and 25.6 versus 20.6 percent on the Bracken (5.1 points, $p = 0.02$). Child demographics show no significant differences (50.9 versus 50.9 percent male; 3.51 versus 3.54 years of age), and neither do parental ethnicity, years of schooling, or age. At the household level, non-graduates score lower on the economic well-being index (7.35 versus 7.68, $p = 0.05$), have lower baseline investment (1.02 versus 1.25 on the factor score, $p < 0.01$), considerably fewer children's books at home (5.21 versus 6.71, $p < 0.01$), and have about half a child more in the household (3.26 versus 2.79, $p < 0.01$).

Table B6: Characteristics of Graduates and Non-Graduates

| | Mean | | Diff. in means |
|---|---|---|---|
| | Graduated | Non Graduated | $p$-value |
| **Test score outcomes** | | | |
| JumpStart baseline score (%) | 53.17 | 44.92 | 0.00 |
| Bracken baseline score (%) | 25.63 | 20.58 | 0.02 |
| **Parental inputs outcomes** | | | |
| No. hours/week read to the child | 1.32 | 0.90 | 0.01 |
| No. of children's books at home | 6.71 | 5.21 | 0.00 |
| No. of skills helped with at home (0–4) | 3.35 | 2.92 | 0.00 |
| No. of hours/day child exposed to TV | 3.11 | 3.00 | 0.71 |
| No. days/month takes child to grocery store | 5.68 | 5.13 | 0.05 |
| Parental inputs factor score | 1.25 | 1.02 | 0.00 |
| **Demographics** | | | |
| Economic well-being index (0–9) | 7.68 | 7.35 | 0.05 |
| Parent years of schooling | 12.39 | 12.38 | 0.99 |
| Parent Hispanic (%) | 58.95 | 61.90 | 0.59 |
| Parent age (years) | 32.83 | 32.18 | 0.41 |
| No. of children in the household | 2.79 | 3.26 | 0.00 |
| Child male (%) | 50.92 | 50.94 | 1.00 |
| Child age (years) | 3.51 | 3.54 | 0.31 |
| Home-school distance (km) | 1.91 | 2.09 | 0.59 |

*Note:* Baseline characteristics of treated families in the experiment sample that completed the JumpStart program (*Graduated*, at least 16 of 22 sessions) and of those that did not (*Non-graduated*), with the $p$-values of their differences. Test scores are in percent of items answered correctly; the economic well-being index runs from 0 to 9 (Appendix B.5).

# Appendix C Additional Analyses on the Impact of JumpStart

## C.1 Detailed Results: The JumpStart Test

Table C1, Panel A, reports intention-to-treat estimates for each of the six JumpStart modules under the four specifications of Section 4, and Appendix B.3 describes the modules. The estimates for sorting and counting, 6.4 and 3.3 percentage points, are not distinguishable from zero.

## C.2 Detailed Results: The Bracken Test

This subsection reports the module-level results behind the Bracken estimates of Section 4.2 (Table 3). Table C1, Panel B, presents intention-to-treat estimates for each of the five modules, and for the two measures derived from the Bracken national norming sample, under the four specifications of Section 4; the instrument and its discontinuation rule are described in Appendix B.3.

Disaggregated by module, the fully controlled estimates are 4.6 percentage points on colors, 2.5 on letters, 2.0 on numbers, 1.5 on sizes, and 0.8 on shapes, and only colors is distinguishable from zero. In the control group the national percentile score is essentially unchanged over the program year, falling from 25.9 to 25.8, and the share classified as delayed or very delayed falls from 51 to 47 percent.

## C.3 Detailed Results: Parental Inputs

This subsection reports the item-level results behind the parental-investment estimates of Section 4.2 (Table 3). Table C1, Panel C, presents intention-to-treat estimates, under the four specifications of Section 4, for each surveyed input: the number of days per week and hours per week parents read to their child, the number of children's books the household owns, whether parents help the child learn the alphabet, colors, numbers, and shapes, the hours per day the television is on when the child is at home, and the frequency of grocery-store trips with the child, while the factor score that aggregates five of them is in Table 3 (Appendix B.4).

Parental inputs increased in the control group between baseline and endline. On average, the number of days someone in the family read to the child rose from 3.1 to 4.1, and the number of children's books at home went from 6.5 to 7.4. Between 81 and 90 percent of control-group parents were already helping their child learn the alphabet, colors, or numbers at baseline, and these figures rose by 7 to 8 percentage points by the end of the year; about two-thirds initially helped with shapes, rising to 83 percent at endline. The factor score rose from 1.20 to 1.42.

## C.4 The Effect of JumpStart Program on Parental Beliefs

This subsection details the belief elicitation and estimates summarized in Section 4.3. We examine one measure of this channel by eliciting parental beliefs from participants in the third cohort.

Table C1: Impact of JumpStart Program (All Cohorts)

| | Control mean | | Intention-to-Treat Effect | | | |
|---|---|---|---|---|---|---|
| | Baseline | Endline | (1) | (2) | (3) | (4) |
| **Panel A: JumpStart test outcomes** | | | | | | |
| Module-specific % score (no. of items) | | | | | | |
| *Colors (10)* | 58.36 | 74.69 | 3.69** | 3.54* | 4.09** | 4.25** |
| | | | (1.73) | (2.11) | (1.96) | (1.92) |
| *Fine motor skills (8)* | 56.95 | 75.65 | 3.84*** | 4.66*** | 4.87*** | 4.82*** |
| | | | (1.29) | (1.53) | (1.48) | (1.48) |
| *Counting (2)* | 40.32 | 62.66 | 3.77 | 4.18 | 3.18 | 3.28 |
| | | | (2.79) | (3.21) | (3.20) | (3.20) |
| *Sorting (1)* | 42.93 | 67.25 | 6.47* | 6.50 | 6.21 | 6.36 |
| | | | (3.25) | (4.17) | (4.05) | (4.04) |
| *Name recognition (1)* | 24.07 | 47.39 | 22.42*** | 24.85*** | 24.56*** | 24.48*** |
| | | | (3.42) | (4.40) | (4.42) | (4.40) |
| *Book handling (4)* | 31.20 | 48.51 | 22.13*** | 21.95*** | 21.39*** | 21.53*** |
| | | | (2.45) | (2.80) | (2.75) | (2.63) |
| **Panel B: Bracken test outcomes** | | | | | | |
| Percentile score | 25.93 | 25.82 | 4.72*** | 4.32** | 2.93* | 3.12* |
| | | | (1.45) | (1.80) | (1.53) | (1.57) |
| % delayed or very delayed | 50.87 | 46.77 | –3.97 | –3.22 | –3.18 | –2.98 |
| | | | (3.15) | (4.01) | (3.93) | (4.13) |
| Module-specific % score (no. of items) | | | | | | |
| *Colors (10)* | 56.43 | 77.39 | 5.27*** | 4.65* | 4.30** | 4.59** |
| | | | (1.84) | (2.33) | (2.12) | (2.19) |
| *Letters (15)* | 20.18 | 28.24 | 4.59** | 4.01* | 2.51 | 2.48 |
| | | | (1.80) | (2.39) | (1.99) | (2.07) |
| *Numbers (18)* | 11.88 | 19.93 | 3.95** | 3.48* | 1.83 | 1.98 |
| | | | (1.57) | (1.95) | (1.66) | (1.64) |
| *Shapes (20)* | 21.60 | 34.91 | 2.09 | 1.41 | 0.77 | 0.84 |
| | | | (1.56) | (1.89) | (1.80) | (1.81) |
| *Sizes (22)* | 20.28 | 29.07 | 1.06 | 1.43 | 1.49 | 1.49 |
| | | | (1.25) | (1.52) | (1.48) | (1.50) |
| **Panel C: Parental inputs outcomes** | | | | | | |
| No. days/week read to the child | 3.06 | 4.10 | 0.66*** | 0.67*** | 0.55*** | 0.56*** |
| | | | (0.16) | (0.20) | (0.19) | (0.18) |
| No. hours/week read to the child | 1.14 | 1.45 | 0.26** | 0.23* | 0.23* | 0.22* |
| | | | (0.11) | (0.13) | (0.12) | (0.12) |
| No. of children's books at home | 6.48 | 7.38 | 1.31*** | 1.22*** | 1.24*** | 1.27*** |
| | | | (0.17) | (0.19) | (0.19) | (0.19) |
| Helps child learn: | | | | | | |
| *Alphabet* | 0.81 | 0.88 | 0.03 | 0.02 | 0.02 | 0.02 |
| | | | (0.02) | (0.03) | (0.03) | (0.03) |
| *Colors* | 0.90 | 0.98 | –0.01 | –0.01 | –0.01 | –0.01 |
| | | | (0.01) | (0.01) | (0.01) | (0.01) |
| *Numbers* | 0.88 | 0.96 | –0.00 | 0.01 | 0.01 | 0.01 |
| | | | (0.02) | (0.02) | (0.02) | (0.02) |
| *Shapes* | 0.68 | 0.83 | –0.04 | –0.04 | –0.04 | –0.05 |
| | | | (0.02) | (0.03) | (0.03) | (0.03) |
| No. of skills helped with at home (0–4) | 3.26 | 3.65 | –0.02 | –0.03 | –0.03 | –0.04 |
| | | | (0.04) | (0.05) | (0.05) | (0.06) |
| No. of hours/day child exposed to TV | 3.24 | 3.61 | –0.23 | –0.23 | –0.20 | –0.16 |
| | | | (0.20) | (0.25) | (0.23) | (0.23) |
| No. days/month takes child to grocery store | 5.66 | 5.76 | 0.06 | 0.07 | 0.08 | 0.08 |
| | | | (0.19) | (0.24) | (0.21) | (0.22) |
| Strata fixed effects | | | | ✓ | ✓ | ✓ |
| Baseline outcome | | | | | ✓ | ✓ |
| Demographic controls | | | | | | ✓ |

*Note:* Intention-to-treat estimates of assignment to JumpStart on the endline module-specific scores of the JumpStart and Bracken tests, with the number of test items in parentheses after each module name, and on the parental inputs, namely days and hours per week someone reads to the child, the number of children's books at home, indicators for helping the child learn the alphabet, colors, numbers and shapes, the number of those four skills helped with at home, hours per day the television is on when the child is home, and grocery trips per month with the child. Two outcomes use the Bracken normalizing national sample: the *percentile score*, in percentiles, and the share of children classified as *delayed or very delayed*, in percentage points. Column (1) has no controls, (2) adds randomization-block indicators, (3) adds the baseline value of the same outcome, and (4) adds baseline demographics. Control-group means of the test scores are in percent correct. Standard errors are clustered at the school-by-cohort level ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Specifically, we asked parents five pairs of belief questions at both baseline and endline. Consistent with the literature, the belief questions present a hypothetical scenario. In our study, the scenario is as follows:

> *Emma and Olivia are two cousins who have just turned three years old. So far, they have been raised in similar ways by their parents. But now that they have turned three years old, their parents are planning to do different things. At Emma's home, the parents are planning to spend 30 minutes per day reading books, teaching Emma about letters, colors, and numbers, teaching them to count things, practicing writing her own name, sorting objects, making balls and snakes with play dough. At Olivia's home, the adults are planning to spend only 10 minutes per day teaching Olivia about the same things.*

Following this scenario, respondents reported their expectations regarding Emma's and Olivia's developmental progress by age four. For each item, respondents could indicate that the child would perform below average, at the average, or above average. We assign a pair-specific score of $+1$ when the respondent expects Emma to achieve a higher developmental level than Olivia, $0$ when the respondent expects the two children to achieve the same level, and $-1$ when the respondent expects Olivia to achieve a higher level than Emma.

We construct two belief measures. *Belief score 1*, reported in the first four columns of Table C2, is the sum of the five pair-specific scores and is defined only for respondents who completed all five pairs. *Belief score 2*, reported in the last four columns, is the sum of the available pair-specific scores for respondents who completed at least one pair. A pair is considered complete only when the respondent provided expectations for both Emma and Olivia. Table C2 reports the estimates, which show no effect of assignment in any specification (Section 4.3). This null result does not rule out changes in other beliefs about the productivity of parental investment.

### C.5 Detailed Results: Heterogeneous Treatment Effects

Figure C1 reports the subgroup estimates of Section 4.4 for the three outcomes, each with the $p$-value of the test that the two subgroup effects are equal. The strata fixed effects and the control coefficients are estimated within each subgroup, and families whose value of the splitting variable is imputed are excluded from that split, so the full-sample estimate need not lie between the two subgroup estimates.

Panel (a) shows the JumpStart test score. Besides the parental-schooling split discussed in Section 4.4, the splits on the two baseline test scores point in the same direction, 10.5 versus 6.2 percentage points by baseline JumpStart score and 9.3 versus 6.1 by baseline Bracken score, but neither difference is statistically significant ($p$-values of 0.11 and 0.27).

Panel (b) shows the Bracken score. Here the sharpest contrast is by parental age, with an estimated effect of 2.5 percentage points for children of older parents and -1.9 percentage points for children of younger parents, a difference that is not statistically significant ($p$-value of 0.10). No

Table C2: Impact of JumpStart on Parental Beliefs Scores (Cohort 3)

| | Belief score 1 | | | | Belief score 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| Treatment dummy | –0.084 | 0.018 | 0.050 | 0.076 | –0.031 | 0.032 | 0.155 | 0.152 |
| *Robust SE* | (0.237) | (0.270) | (0.291) | (0.287) | (0.231) | (0.256) | (0.257) | (0.249) |
| *School level SE* | (0.252) | (0.303) | (0.366) | (0.371) | (0.238) | (0.282) | (0.308) | (0.303) |
| Baseline belief score | | | 0.337* | 0.323* | | | 0.350** | 0.330* |
| *Robust SE* | | | (0.125) | (0.130) | | | (0.115) | (0.124) |
| *School level SE* | | | (0.165) | (0.172) | | | (0.151) | (0.166) |
| Constant | 3.836*** | 3.783*** | 2.438*** | –1.283 | 3.781*** | 3.747*** | 2.312*** | –2.372 |
| *Robust SE* | (0.158) | (0.180) | (0.533) | (4.315) | (0.163) | (0.174) | (0.502) | (3.996) |
| *School level SE* | (0.147) | (0.157) | (0.649) | (5.833) | (0.150) | (0.150) | (0.600) | (5.157) |
| $R^2$ | 0.000 | 0.188 | 0.263 | 0.320 | 0.000 | 0.186 | 0.249 | 0.303 |
| Observations | 279 | 279 | 256 | 256 | 293 | 293 | 286 | 286 |
| Strata fixed effects | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Baseline outcome | | | ✓ | ✓ | | | ✓ | ✓ |
| Demographic controls | | | | ✓ | | | | ✓ |

*Note:* Intention-to-treat estimates of assignment to JumpStart on endline parental belief scores, third cohort only. The scores are constructed from five pairs of vignette questions. *Belief score 1* is the sum of all five pair-specific scores and is defined only for families that answered all five pairs. *Belief score 2* is the sum of the available pair-specific scores for families that answered at least one pair. Column (1) includes only the treatment indicator; Column (2) adds randomization-block indicators; Column (3) adds the corresponding baseline belief score; and Column (4) additionally includes demographic controls. The table reports robust and school-clustered standard errors. Significance stars are based on school-clustered $p$-values: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

other split comes close to significance, and in particular the education gradient that appears on the JumpStart test is muted on the Bracken (2.6 versus 1.1 percentage points, $p$-value of 0.67).

Panel (c) turns to the parental-investment factor score, the summary measure of the home learning environment. Every subgroup estimate is positive, and they range from 0.03 to 0.34 points around a full-sample effect of 0.26. None of the six splits is close to significant, the smallest $p$-value being 0.31, for the widest raw gap, 0.24 for children of above-median-educated parents against 0.03 for the rest, but the effect is estimated imprecisely in the smaller below-median-schooling cell, so the data cannot distinguish a genuine contrast from noise. We therefore detect no heterogeneity in the program's effect on the overall home learning environment, even where its individual components move differently.

## C.6 Detailed Results: Attrition

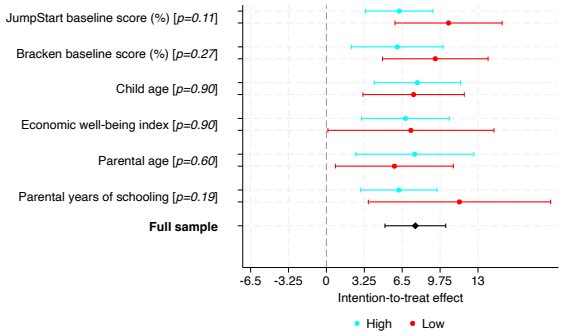
This subsection details the attrition analysis of Section 4.5. Appendix Table C3 compares attriters with stayers in the full sample and separately by treatment status.

We evaluate the sensitivity of the ITT estimates using inverse probability weighting, the Heckman selection model, and Lee bounds. We additionally adopt a procedure designed to address unbalanced attrition, implemented in four steps:

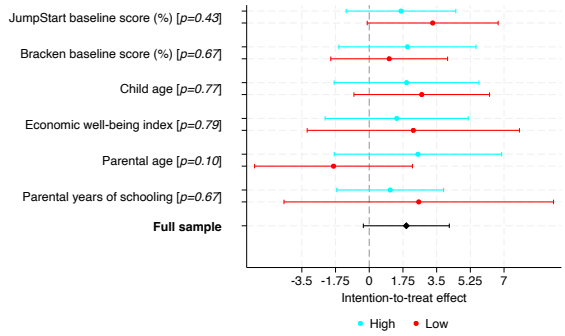
1. **Estimate the probability of attrition.** Let $D = 0$ indicate an individual who remains in the

Figure C1: Heterogeneous Intention-to-Treat Effects

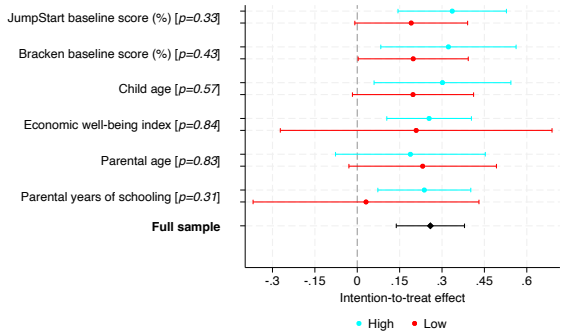
(a) JumpStart Test




(b) Bracken




(c) Parental Inputs Factor Score




*Note:* Subgroup intention-to-treat estimates on the endline JumpStart score (a), the Bracken score (b), and the parental-investment factor score (c). Each pair of rows splits the experiment sample by a baseline characteristic, above versus below the sample median, and re-estimates column (4) of Table 3 within each subgroup. Test-score effects are in percentage points; the factor score is in weekly reading hours. Bars are 95% confidence intervals from standard errors clustered at the school-by-cohort level; the bracketed $p$-value tests the equality of the two subgroup effects, from a regression that interacts every control and the randomization-block fixed effects with the subgroup indicator. Families whose value of the splitting variable is imputed are excluded from that split. The bottom row, in black, is the full-sample estimate.

sample, and let $D = 1$ indicate an individual who attrites. We estimate a probit model of $D$ on demographic characteristics, cohort dummies, and the baseline JumpStart test score.[42]

2. **Create matched pairs.** Using the estimated probabilities of attrition, we match each parent in the treatment group, without replacement, with the control-group parent whose probability of attrition is closest. The matching is performed within each stratum (a school $\times$ blocking group $\times$ year cell, the level at which assignment was actually randomized), ensuring that every matched pair consists of one treatment and one control parent.[43]

[42] See Appendix Table C4 for these first-stage results. Attrition is lower among Hispanic parents, among older and more educated parents, and in the second and third cohorts. The baseline JumpStart test score does not predict attrition.

[43] Restricting matches to the year in which the lottery was run matters here because attrition varies sharply across the three cohorts, from 29 percent in the first year to 15 percent in the third, and because the blocking cut points were computed separately within each year's lottery sample.

Table C3: Characteristics of JumpStart Program Attrition

| | Full sample | | | Treatment group | | | Control group | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | | | Mean | | | Mean | | |
| | Attriters | Non-attriters | $p$-value | Attriters | Non-attriters | $p$-value | Attriters | Non-attriters | $p$-value |
| **Test score outcomes** | | | | | | | | | |
| JumpStart baseline score (%) | 51.96 | 50.96 | 0.59 | 51.37 | 51.37 | 1.00 | 52.30 | 50.45 | 0.46 |
| Bracken baseline score (%) | 26.76 | 23.86 | 0.04 | 27.52 | 24.53 | 0.20 | 26.32 | 23.05 | 0.06 |
| **Parental inputs outcomes** | | | | | | | | | |
| No. hours/week read to the child | 1.29 | 1.19 | 0.37 | 1.20 | 1.23 | 0.86 | 1.34 | 1.14 | 0.17 |
| No. of children's books at home | 6.40 | 6.43 | 0.93 | 6.44 | 6.38 | 0.88 | 6.38 | 6.48 | 0.76 |
| No. of skills helped with at home (0–4) | 3.25 | 3.26 | 0.89 | 3.16 | 3.26 | 0.46 | 3.30 | 3.26 | 0.72 |
| No. of hours/day child exposed to TV | 3.20 | 3.15 | 0.82 | 3.33 | 3.09 | 0.42 | 3.12 | 3.24 | 0.65 |
| No. days/month takes child to grocery store | 5.60 | 5.61 | 0.96 | 5.65 | 5.56 | 0.77 | 5.56 | 5.66 | 0.70 |
| Parental inputs factor score | 1.22 | 1.20 | 0.55 | 1.20 | 1.20 | 0.96 | 1.23 | 1.20 | 0.45 |
| **Demographics** | | | | | | | | | |
| Economic well-being index (0–9) | 7.40 | 7.62 | 0.07 | 7.31 | 7.61 | 0.11 | 7.46 | 7.63 | 0.26 |
| Parent years of schooling | 11.83 | 12.24 | 0.11 | 11.52 | 12.38 | 0.03 | 12.01 | 12.06 | 0.87 |
| Parent Hispanic (%) | 49.36 | 60.81 | 0.00 | 43.68 | 59.59 | 0.01 | 52.70 | 62.28 | 0.04 |
| Parent age (years) | 31.41 | 32.60 | 0.02 | 31.79 | 32.69 | 0.27 | 31.18 | 32.50 | 0.05 |
| No. of children in the household | 2.81 | 2.85 | 0.62 | 2.95 | 2.89 | 0.66 | 2.72 | 2.81 | 0.44 |
| Child male (%) | 47.03 | 49.44 | 0.51 | 35.63 | 50.92 | 0.01 | 53.69 | 47.64 | 0.21 |
| Child age (years) | 3.55 | 3.52 | 0.21 | 3.58 | 3.52 | 0.05 | 3.52 | 3.52 | 0.89 |
| Home-school distance (km) | 2.00 | 1.83 | 0.35 | 2.23 | 1.95 | 0.42 | 1.87 | 1.70 | 0.28 |

*Note:* Baseline characteristics of families lost between baseline and endline (*Attriters*) and of those observed at endline (*Non-attriters*), in the full sample and separately for the treatment and control groups. The $p$-values test equality of the two means.

3. **Refine the sample using endline data.** If either member of a matched pair is missing from the endline dataset, we exclude the entire pair; if a control parent is missing at endline, the matched treatment parent is also removed. This yields a smaller but balanced sample, eliminating differential attrition between treatment and control groups.

4. **Estimate ITT parameters.** We re-estimate the ITT effects on this refined sample. In Table C5, we refer to this procedure as "Balancing."

Table C5 presents the results for the JumpStart and Bracken scores and the parental-inputs factor score that the model of Section 5 uses as its investment measure. The full set of module- and item-level results is in Appendix Table C6, Panels A to C.

Table C4: First-stage Results of Probit Estimation of Attrition

| | Estimate | SE |
|---|---|---|
| JumpStart baseline score | -0.010 | (0.179) |
| Child male | -0.005 | (0.080) |
| Child age (years) | 0.058 | (0.144) |
| Parent Hispanic | -0.409*** | (0.107) |
| Parent years of schooling | -0.035** | (0.015) |
| Parent age (years) | -0.019* | (0.011) |
| No. of children in the household | -0.033 | (0.041) |
| Economic well-being index (0–9) | -0.021 | (0.031) |
| Cohort 2 | -0.353*** | (0.128) |
| Cohort 3 | -0.437*** | (0.135) |
| Missing indicator | 0.193 | (0.123) |
| Constant | 0.699 | (0.713) |
| Observations | 1,126 | |
| Pseudo R-squared | 0.046 | |

*Note:* First-stage probit used by the inverse-probability-weighting, balancing and Heckman corrections of Table C5. *Attrit* equals one for families observed at baseline but not at endline. Regressors are the demographic controls of the main ITT specification, cohort dummies (first cohort omitted), and the baseline JumpStart score; missing controls are mean-imputed and flagged by the missing indicator. Standard errors clustered at the school-by-cohort level ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Table C5: Sensitivity of the Intention-to-Treat Effect with Respect to Attrition (Main Outcomes)

| | Point estimates | | | | Lee bounds | |
|---|---|---|---|---|---|---|
| | No correction | IPW | Balancing | Heckman selection | Lower bound | Upper bound |
| JumpStart test score (% correct) | 7.63*** | 7.58*** | 7.10*** | 7.63*** | 3.57*** | 10.22*** |
| *Robust SE* | (1.11) | (1.12) | (1.20) | (1.11) | (1.02) | (1.01) |
| *School level SE* | (1.64) | (1.59) | (1.69) | (1.64) | (1.21) | (1.39) |
| *School-cohort level SE* | (1.33) | (1.29) | (1.55) | (1.33) | (1.12) | (1.14) |
| *N* | 890 | 890 | 610 | 890 | 1,126 | 1,126 |
| Bracken test score (% correct) | 1.93* | 1.99* | 1.48 | 1.95* | -1.01 | 4.19*** |
| *Robust SE* | (0.90) | (0.92) | (0.96) | (0.90) | (0.77) | (0.80) |
| *School level SE* | (1.13) | (1.15) | (1.23) | (1.12) | (0.91) | (0.97) |
| *School-cohort level SE* | (1.14) | (1.13) | (1.30) | (1.14) | (0.99) | (0.91) |
| *N* | 888 | 888 | 610 | 888 | 1,126 | 1,126 |
| Parental inputs factor score | 0.26*** | 0.26*** | 0.19** | 0.26*** | 0.07 | 0.38*** |
| *Robust SE* | (0.06) | (0.06) | (0.07) | (0.06) | (0.06) | (0.06) |
| *School level SE* | (0.07) | (0.07) | (0.08) | (0.07) | (0.04) | (0.08) |
| *School-cohort level SE* | (0.06) | (0.06) | (0.07) | (0.06) | (0.06) | (0.06) |
| *N* | 852 | 852 | 593 | 852 | 1,126 | 1,126 |

*Note:* This table shows the sensitivity of intention-to-treat (ITT) treatment effect parameters accounting for attrition. The first column uses the specification with randomization-strata fixed effects, the baseline outcome, and demographic controls (column (4) of Table 3). The second, third, and fourth columns report inverse-probability weighting, the balancing procedure, and the Heckman selection model. The last two columns report Lee sharp bounds. The first two outcomes are measured in percentage points; the parental-inputs factor score is expressed in weekly reading hours. We report heteroskedasticity-robust, school-clustered, and school-by-cohort-clustered standard errors; the Lee-bound standard errors are bootstrap based (500 replications, resampling at the corresponding cluster level). Stars refer to the school-by-cohort clustered $p$-value ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Table C6: Sensitivity of the Intention-to-Treat Effect with Respect to Attrition (All Outcomes)

| | Point estimates | | | | Lee bounds | |
|---|---|---|---|---|---|---|
| | No correction | IPW | Balancing | Heckman selection | Lower bound | Upper bound |
| **Panel A: JumpStart test outcomes** | | | | | | |
| Module-specific % score (no. of items) | | | | | | |
| *Colors (10)* | 4.25** | 4.24** | 2.70 | 4.25** | -1.01 | 8.92*** |
| | (1.92) | (1.90) | (2.28) | (1.92) | (1.70) | (1.60) |
| *Fine motor skills (8)* | 4.82*** | 4.77*** | 4.86*** | 4.82*** | 0.48 | 8.28*** |
| | (1.48) | (1.46) | (1.73) | (1.49) | (1.18) | (1.30) |
| *Counting (2)* | 3.28 | 3.15 | 1.25 | 3.34 | -4.96* | 11.54*** |
| | (3.20) | (3.23) | (3.54) | (3.27) | (2.79) | (2.78) |
| *Sorting (1)* | 6.36 | 6.76 | 4.80 | 6.30 | -2.66 | 15.77*** |
| | (4.04) | (4.11) | (4.42) | (4.03) | (3.21) | (3.42) |
| *Name recognition (1)* | 24.48*** | 24.36*** | 23.10*** | 24.49*** | 13.12*** | 31.48*** |
| | (4.40) | (4.28) | (5.00) | (4.40) | (3.41) | (3.66) |
| *Book handling (4)* | 21.53*** | 21.41*** | 23.59*** | 21.51*** | 12.18*** | 26.36*** |
| | (2.63) | (2.61) | (3.10) | (2.62) | (2.34) | (2.24) |
| **Panel B: Bracken test outcomes** | | | | | | |
| Percentile score | 3.12* | 3.15** | 2.40 | 3.17** | -1.40 | 6.26*** |
| | (1.57) | (1.58) | (1.72) | (1.57) | (1.36) | (1.34) |
| % delayed or very delayed | -2.98 | -3.10 | -1.23 | -3.01 | -11.30*** | 6.32* |
| | (4.13) | (4.07) | (4.72) | (4.14) | (3.32) | (3.70) |
| Module-specific % score (no. of items) | | | | | | |
| *Colors (10)* | 4.59** | 4.45** | 3.87 | 4.59** | -1.21 | 10.14*** |
| | (2.19) | (2.19) | (2.63) | (2.19) | (1.83) | (1.65) |
| *Letters (15)* | 2.48 | 2.51 | 1.69 | 2.49 | -3.33** | 6.61*** |
| | (2.07) | (2.09) | (2.38) | (2.05) | (1.69) | (1.66) |
| *Numbers (18)* | 1.98 | 2.04 | 1.08 | 1.99 | -2.88** | 5.74*** |
| | (1.64) | (1.66) | (1.84) | (1.64) | (1.36) | (1.45) |
| *Shapes (20)* | 0.84 | 0.95 | -0.21 | 0.87 | -3.14** | 4.27*** |
| | (1.81) | (1.81) | (1.89) | (1.81) | (1.49) | (1.54) |
| *Sizes (22)* | 1.49 | 1.56 | 2.52 | 1.51 | -2.12* | 4.23*** |
| | (1.50) | (1.46) | (1.76) | (1.52) | (1.24) | (1.18) |
| **Panel C: Parental inputs outcomes** | | | | | | |
| No. days/week read to the child | 0.56*** | 0.57*** | 0.53** | 0.58*** | 0.07 | 0.94*** |
| | (0.18) | (0.19) | (0.22) | (0.18) | (0.16) | (0.17) |
| No. hours/week read to the child | 0.22* | 0.23* | 0.22 | 0.22* | -0.13 | 0.44*** |
| | (0.12) | (0.13) | (0.16) | (0.13) | (0.09) | (0.12) |
| No. of children's books at home | 1.27*** | 1.24*** | 1.27*** | 1.27*** | 0.66*** | 1.58*** |
| | (0.19) | (0.19) | (0.21) | (0.19) | (0.15) | (0.16) |
| Helps child learn: | | | | | | |
| *Alphabet* | 0.02 | 0.02 | 0.00 | 0.02 | -0.04 | 0.09*** |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.02) |
| *Colors* | -0.01 | -0.01 | -0.02 | -0.01 | -0.03*** | 0.02*** |
| | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) |
| *Numbers* | 0.01 | 0.01 | 0.00 | 0.01 | -0.02* | 0.04*** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.01) | (0.01) |
| *Shapes* | -0.05 | -0.04 | -0.09** | -0.05 | -0.11*** | 0.06** |
| | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.03) |
| No. of skills helped with at home (0–4) | -0.04 | -0.04 | -0.10 | -0.04 | -0.14*** | 0.12*** |
| | (0.06) | (0.06) | (0.07) | (0.06) | (0.05) | (0.04) |
| No. of hours/day child exposed to TV | -0.16 | -0.14 | -0.09 | -0.16 | -0.68*** | 0.30 |
| | (0.23) | (0.24) | (0.28) | (0.23) | (0.19) | (0.21) |
| No. days/month takes child to grocery store | 0.08 | 0.11 | 0.01 | 0.07 | -0.36* | 0.55*** |
| | (0.22) | (0.22) | (0.23) | (0.21) | (0.19) | (0.18) |

*Note:* This table reports the attrition-sensitivity analyses for the module-specific JumpStart and Bracken test outcomes and the individual parental-input outcomes. The columns apply the same corrections as Table C5. Standard errors are clustered at the school-by-cohort level; the Lee-bound standard errors are bootstrap based on 500 replications at that level. Stars refer to the school-by-cohort clustered $p$-value ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

# Appendix D Structural Model: Derivations and Identification

This section derives the estimating equations of Section 5.4 and shows how the structural parameters are recovered from them. We use the notation of Sections 5.1.2–5.2.3: the technology (3), the utility function (4), the taste and target specifications (5), and the vector $\boldsymbol{Z}_i$, which contains a constant and each distinct observed covariate in $X_{i0}$, $\boldsymbol{Z}_{X,i}$, $\boldsymbol{Z}_{K,i}$, $\boldsymbol{Z}_{G,i}$, and $\boldsymbol{W}_i$. Assumptions 1 and 2 of Section 5.4 are maintained throughout.

## D.1 Optimal Investment

Let $\gamma_{1i}(g)$ be the marginal taste in (5) under graduation status $g$, and let $b_i(g) \equiv \gamma_{2i}+\kappa\beta_{X_1}+\gamma_{1i}(g)$ be the unconstrained investment choice. If investment cannot be negative, strict concavity gives

$$X_{i1}^{*}(g) = \max\{0, b_i(g)\}.$$

The choice is interior when $b_i(g) > 0$ and is at zero otherwise. An interior solution is characterized by the first-order condition

$$\kappa\beta_{X_1} + \gamma_{1i}(g) - \left(X_{i1} - \gamma_{2i}\right) = 0,$$

which rearranges to $X_{i1}^{*} = \kappa\beta_{X_1} + \gamma_{1i}(g) + \gamma_{2i}$, equation (6) of the text. Substituting the taste and target specifications (5) and collecting intercepts into $\widetilde{\gamma}_0$ as in (8) yields (7), and (9) for households with $G_i = 0$. This condition depends on the full expression for $b_i(g)$, regardless of the sign of $\kappa$. Investment equal to the target $\gamma_{2i} > 0$ is also an interior choice, which occurs when $\kappa\beta_{X_1}+\gamma_{1i}(g) = 0$.

The linear investment rule in equation (7) applies when both $b_i(0)$ and $b_i(1)$ are positive. Investment then changes by $\gamma_G$ on graduation. If either choice is at zero, the change is $\max\{0, b_i(0) + \gamma_G\} - \max\{0, b_i(0)\}$ and need not equal $\gamma_G$. The graduation equation and investment-channel formula would then need to account for the corner solution.

We focus on the interior specification as an assumption. The data cannot test this assumption directly. The origin of the investment factor score is set by the normalization of Appendix B.4, so the sign of a score carries no information about a corner, and no family reveals the investment it would choose under the other graduation status. The items from which the score is built nonetheless suggest that corner solutions are rare, at least under the status each family actually has. Among the 887 families with all three items recorded at endline, 2.1 percent report never reading to the child, 1.7 percent report having no children's books, and 0.3 percent report helping with none of the four school-readiness skills, while a single family reports all three.

## D.2 The Graduation Decision

A parent in the treatment group graduates when the utility of graduating, evaluated at the investment that is optimal under graduation, weakly exceeds the utility of not graduating, eval-

uated at the investment that is optimal without it. Substituting (7) and (9) into (4), expanding the quadratic loss, and collecting the terms that multiply $G_i$ delivers the index (10) with error $\omega_i = \kappa\varepsilon_{i1} + \gamma_G\nu_i - \eta_i$, the composite slopes of (11), and intercept

$$\widetilde{\alpha}_0 = \alpha_0 + \kappa\beta_G + \gamma_G\big(\gamma_{10} + \gamma_{20} + \kappa\beta_{X_1}\big) + \tfrac{1}{2}\gamma_G^2 = \alpha_0 + \kappa\beta_G + \gamma_G\widetilde{\gamma}_0 + \tfrac{1}{2}\gamma_G^2. \tag{35}$$

The four terms in (35) have distinct interpretations. They are the intrinsic taste for completing the program, the value of the direct knowledge gain, the value of the higher marginal taste for investment at the investment level the parent would choose without graduating, and the gain from re-optimizing investment in response to that higher taste. The last term is nonnegative, and positive whenever $\gamma_G \neq 0$. It is the utility a graduating parent gains by moving investment from the level chosen without graduating to the new optimum, and a move to an optimum can never lower utility.

### D.3 The Reduced-Form Knowledge Equation

Substituting optimal investment (7) into the technology (3) and collecting terms gives (12), with composite error $\xi_i = \beta_{X_1}\nu_i + \varepsilon_i$ and the coefficient map (13). The same expression applies to households with $G_i = 0$, in which case the graduation term vanishes.

### D.4 Identification of the Reduced Form

**The graduation probability.** Under Assumption 2, $\omega_i \sim N(0,\sigma_\omega^2)$ and is independent of $\boldsymbol{Z}_i$, and under Assumption 1 it is independent of $T_i$ as well. For a household with $T_i = 1$,

$$\Pr(G_i = 1 \mid T_i = 1,\boldsymbol{Z}_i) = \Phi\left(\frac{\widetilde{\alpha}_0 + \widetilde{\alpha}_{X_0}X_{i0} + \widetilde{\boldsymbol{\alpha}}'_{Z_X}\boldsymbol{Z}_{X,i} + \boldsymbol{\alpha}'_{Z_G}\boldsymbol{Z}_{G,i} + \widetilde{\boldsymbol{\alpha}}'_W\boldsymbol{W}_i}{\sigma_\omega}\right) = \Phi(\boldsymbol{a}'\boldsymbol{Z}_i),$$

so the probit identifies the coefficient on each distinct covariate:

$$a_0 = \frac{\widetilde{\alpha}_0}{\sigma_\omega}, \qquad a_{X_0} = \frac{\widetilde{\alpha}_{X_0}}{\sigma_\omega}, \qquad a_j = \frac{\widetilde{\alpha}_{Z_X,j} + \alpha_{Z_G,j} + \widetilde{\alpha}_{W,j}}{\sigma_\omega}. \tag{36}$$

The last expression covers every other distinct covariate in $\boldsymbol{Z}_i$, with a zero contribution from each payoff term in which it is absent. A covariate appearing only in $\boldsymbol{Z}_{K,i}$ therefore has $a_j = 0$. Households in the control group have $\Pr(G_i = 1 \mid T_i = 0,\boldsymbol{Z}_i) = 0$ and are uninformative about $\boldsymbol{a}$.

**The investment equation.** Selection into graduation truncates the distribution of $\nu_i$, which enters $\omega_i$ through $\gamma_G\nu_i$. Taking the expectation of (7) conditional on $(\boldsymbol{Z}_i,G_i)$ for a household in the treatment group gives

$$\mathbb{E}\big[X_{i1} \mid T_i = 1,G_i,\boldsymbol{Z}_i\big] = \widetilde{\gamma}_0 + \gamma_{X_0}X_{i0} + \gamma_G G_i + \boldsymbol{\gamma}'_{Z_X}\boldsymbol{Z}_{X,i} + \rho_\nu\lambda_i, \tag{37}$$

with $\lambda_i$ as in (20) and $\rho_\nu = \mathrm{Cov}(\nu_i,\omega_i)/\sigma_\omega$. Equation (37) is linear in $(X_{i0},G_i,\boldsymbol{Z}_{X,i},\lambda_i)$, so $(\widetilde{\gamma}_0,\gamma_{X_0},\gamma_G,\boldsymbol{\gamma}_{Z_X},\rho_\nu)$ are identified on the treated subsample. For control households $\mathbb{E}[\nu_i \mid T_i = 0,\boldsymbol{Z}_i] = 0$, the correction term drops out, and $(\widetilde{\gamma}_0,\gamma_{X_0},\boldsymbol{\gamma}_{Z_X})$ are identified there as well. Pooling the two arms therefore over-identifies these parameters.

**The knowledge equation.** Applying the same argument to (12),

$$\begin{aligned}\mathbb{E}\big[K_{i1} \mid T_i = 1,G_i,K_{i0},\boldsymbol{Z}_i\big] = \widetilde{\beta}_0 + \beta_{K_0}K_{i0} + \widetilde{\beta}_{X_0}X_{i0} + \big(\widetilde{\beta}_G + \boldsymbol{\beta}_W'\boldsymbol{W}_i\big)G_i \\ + \widetilde{\boldsymbol{\beta}}_{X,K}'\boldsymbol{Z}_{X,K,i} + \rho_0(1-G_i)\lambda_{0i} + \rho_1 G_i\lambda_{1i},\end{aligned} \tag{38}$$

where

$$\rho_0 = \frac{\mathrm{Cov}(\xi_i,\omega_i)}{\sigma_\omega}, \qquad \rho_1 = \frac{\mathrm{Cov}(\xi_i+\varepsilon_{i1},\omega_i)}{\sigma_\omega}.$$

Two distinct correction terms are required because graduates and non-graduates differ not only in the truncation of $\xi_i$ but also in whether the idiosyncratic gain $\varepsilon_{i1}$ is realized in the outcome. Their difference is precisely the selection wedge (25). Linearity of (38) identifies

$$\big(\widetilde{\beta}_0,\ \beta_{K_0},\ \widetilde{\beta}_{X_0},\ \widetilde{\beta}_G,\ \boldsymbol{\beta}_W,\ \widetilde{\boldsymbol{\beta}}_{X,K},\ \rho_0,\ \rho_1\big).$$

### D.5 Recovering the Structural Parameters

**The technology and the direct effect.** The coefficient map (13) contains $\beta_{X_1}$ in several places while the investment equation contains the corresponding preference parameters alone. The baseline-investment ratio (18) identifies $\beta_{X_1}$ when $\gamma_{X_0} \neq 0$. A covariate in $\boldsymbol{Z}_{X,i}$ but not $\boldsymbol{Z}_{K,i}$ could provide an additional ratio; a shared control cannot, because its reduced-form knowledge coefficient also contains its technology effect. Given $\beta_{X_1}$, the remaining production parameters follow as

$$\beta_G = \widetilde{\beta}_G - \beta_{X_1}\gamma_G, \qquad \beta_0 = \widetilde{\beta}_0 - \beta_{X_1}\widetilde{\gamma}_0.$$

The coefficients $\beta_{K_0}$ and $\boldsymbol{\beta}_W$ are read directly from (38). For each covariate in $\boldsymbol{Z}_{K,i}$, its technology coefficient is recovered from $\widetilde{\boldsymbol{\beta}}_{X,K}$ by subtracting $\beta_{X_1}$ times its investment-rule coefficient when that covariate also appears in $\boldsymbol{Z}_{X,i}$. The treatment-effect formulas (22)–(24) do not require these technology-control coefficients.

**The scale of the selection error.** The probit identifies the index only up to $\sigma_\omega$. The investment equation supplies the missing units. Since $\widetilde{\alpha}_{X_0} = \gamma_{X_0}\gamma_G$ is identified from (37) while $a_{X_0} = \widetilde{\alpha}_{X_0}/\sigma_\omega$ is identified from (36),

$$\sigma_\omega = \frac{\gamma_{X_0}\gamma_G}{a_{X_0}}, \tag{39}$$

which requires a nonzero baseline-investment coefficient in both equations and that $X_{i0}$ does not also enter the other graduation-payoff covariate vectors. Intuitively, the probit reveals how much

a shift in baseline investment moves the graduation decision, and the investment equation reveals how large that shift is in units of investment; the ratio converts utility into observable units. The utility intercept then follows as

$$\alpha_0 = \sigma_\omega a_0 - \kappa\beta_G - \gamma_G\widetilde{\gamma}_0 - \tfrac{1}{2}\gamma_G^2.$$

The coefficients in $\boldsymbol{\alpha}_{Z_G}$ are recovered from the corresponding coefficients of $\sigma_\omega \boldsymbol{a}$ after subtracting any other graduation-payoff contributions to the same covariates.

**The marginal utility of knowledge.** When the return to graduation is allowed to vary with observables, $\kappa$ is separately identified. For a component of $\boldsymbol{W}_i$ that does not also enter the other graduation-payoff covariate vectors, $a_{W,j} = \widetilde{\alpha}_{W,j}/\sigma_\omega$. From $\widetilde{\boldsymbol{\alpha}}_W = \kappa\boldsymbol{\beta}_W$ and (36),

$$\kappa = \frac{\sigma_\omega a_{W,j}}{\beta_{W,j}} = \frac{a_{W,j}}{a_{X_0}} \cdot \frac{\gamma_{X_0}\gamma_G}{\beta_{W,j}} \quad \text{for any } j,$$

in which $\sigma_\omega$ cancels. The logic is that $\boldsymbol{W}_i$ shifts the graduation decision only by shifting the anticipated knowledge gain; comparing how much it moves the decision with how much it moves the gain recovers the utility value of that gain. Without $\boldsymbol{W}_i$ the two are indistinguishable and $\kappa$ is absorbed into $\sigma_\omega$. Tables 4 and D1 report negative estimates of $\beta_{W,K_{i0}\geq\text{med}}$ and $\beta_{W,\text{educ}}$ (lower returns to graduation for children with higher baseline score and more educated parents) and positive but imprecisely estimated $a_{W,K_0}$ and $a_{W,\text{educ}}$, leading to imprecise estimates of $\kappa$ with negative sign.

**Composites and non-identified parameters.** The taste and target intercepts enter only through their sum,

$$\gamma_{10} + \gamma_{20} = \widetilde{\gamma}_0 - \kappa\beta_{X_1},$$

and are not separately identified. The error covariances are recovered only in the combinations $\text{Cov}(\nu_i,\omega_i) = \rho_\nu\sigma_\omega$, $\text{Cov}(\xi_i,\omega_i) = \rho_0\sigma_\omega$, and $\text{Cov}(\xi_i + \varepsilon_{i1},\omega_i) = \rho_1\sigma_\omega$; the individual entries of the covariance matrix, and in particular any covariance involving the taste shock $\eta_i$, are not separately identified without further restrictions. Finally, if $\gamma_G = 0$ the production parameters are unaffected, but (39) fails, and $(\kappa,\alpha_0,\boldsymbol{\alpha}_{Z_G})$ are identified only up to scale.

### D.6 Treatment Effects

Differencing (3) at $G_i = 1$ and $G_i = 0$ and substituting the optimal rules (7) and (9) gives (21): investment changes by exactly $\gamma_G$, which enters knowledge at rate $\beta_{X_1}$ and produces the investment channel $\beta_{X_1}\gamma_G$, while the direct channel is $\beta_G + \boldsymbol{\beta}_W'\boldsymbol{W}_i$ and the unobserved gain is $\varepsilon_{i1}$. Averaging over the population gives (22), since $\mathbb{E}[\varepsilon_{i1}] = 0$ by normalization. Averaging over graduates and non-graduates within the treated arm requires the truncated means of the idiosyncratic gain,

$$\mathbb{E}\big[\varepsilon_{i1} \mid T_i = 1,G_i = 1\big] = (\rho_1 - \rho_0)\,\mathbb{E}\big[\lambda_{1i} \mid T_i = 1,G_i = 1\big],$$

and the analogous expression with $\lambda_{0i}$ for non-graduates, which yield (23) and (24). When $\rho_1 = \rho_0$ the idiosyncratic gain is uncorrelated with the graduation decision, the truncation terms vanish, and the three treatment effects differ only through the composition of $\boldsymbol{W}_i$.

### D.7 The Model-Implied MTE and Treatment-Effect Parameters

All treatment-effect parameters are integrals of the MTE in (30) against different weights on the resistance quantile $u$ (Heckman and Vytlacil, 2005). The single analytical ingredient is the truncated-mean integral of the standard normal quantile function. Substituting $u = \Phi(v)$, $du = \phi(v)\,dv$, and using $\frac{d}{dv}\big[-\phi(v)\big] = v\,\phi(v)$, gives

$$\int_0^p \Phi^{-1}(u)\,du \;=\; \int_{-\infty}^{\Phi^{-1}(p)} v\,\phi(v)\,dv \;=\; -\,\phi\big(\Phi^{-1}(p)\big), \qquad p \in [0,1].$$

Setting $p = 1$ gives $\int_0^1 \Phi^{-1}(u)\,du = 0$ (since $\phi(v) \to 0$ as $v \to \infty$), and therefore $\int_p^1 \Phi^{-1}(u)\,du = +\,\phi(\Phi^{-1}(p))$. Throughout, write $s \equiv \rho_1 - \rho_0$ for the selection wedge and $h_i \equiv \boldsymbol{a}'\boldsymbol{Z}_i = \Phi^{-1}(p_i)$ for the index of the probit (14), so $\mathrm{MTE}(u) = \widetilde{\beta}_G - s\,\Phi^{-1}(u)$, and recall $G_i = T_i\mathbf{1}\{U_{G,i} \le p_i\}$ with $U_{G,i} \sim \mathrm{Uniform}[0,1]$ independent of $\boldsymbol{Z}_i$.

With observed heterogeneity in the return to graduation, every level term below shifts by $\boldsymbol{\beta}_W'\boldsymbol{W}_i$ and the derivations are unchanged conditional on $\boldsymbol{W}_i$. Averaging over the whole unit interval gives $\mathrm{ATE} = \widetilde{\beta}_G$, since $\int_0^1 \Phi^{-1}(u)\,du = 0$. Averaging over $u \le p_i$ gives the effect for a graduate with propensity $p_i$, $\widetilde{\beta}_G + s\,\phi(h_i)/\Phi(h_i) = \widetilde{\beta}_G + s\,\lambda_{1i}$, and averaging over $u > p_i$ gives the effect for a never-taker, $\widetilde{\beta}_G - s\,\phi(h_i)/(1-\Phi(h_i)) = \widetilde{\beta}_G + s\,\lambda_{0i}$. Averaging these over graduates and over offered non-graduates gives (23) and (24). The household-level ITT is $p_i$ times its effect on the treated, $p_i\widetilde{\beta}_G + s\,\phi(h_i)$, and its average over the offered arm equals $\pi \cdot \mathrm{ATT}$. Finally, the average of the MTE over a window $[a,b]$, the object compared with the bounds in Appendix F.2, is $\widetilde{\beta}_G - s\,[\phi(\Phi^{-1}(a)) - \phi(\Phi^{-1}(b))]/(b-a)$, with $\phi(\Phi^{-1}(0)) = \phi(\Phi^{-1}(1)) = 0$ understood as limits.

### D.8 Structural Model Estimates

Appendix Table D1 reports the remaining coefficients of the two specifications of Table 4. Appendix Table D2 reports estimates of the investment equation and the knowledge production function on the control arm alone, which require no correction for selection. There the marginal product of investment is 1.7, with a standard error of 2.9, against 6.1 when both arms are pooled, and the control-arm confidence interval contains the pooled estimate. The other control-arm estimates are close to their pooled counterparts, in particular the inertia of investment (0.88 against 0.81) and the persistence of knowledge (0.38 against 0.36), and most covariate coefficients lie within one standard error of the pooled ones.

Table D1: Structural Model Estimates (contd.)

| | | (1) No $G_i \times W_i$ interactions | | | (2) With $G_i \times W_i$ interactions | | |
|---|---|---|---|---|---|---|---|
| Parameter | Description | Estimate | SE (delta) | SE (boot.) | Estimate | SE (delta) | SE (boot.) |
| *Panel A: Graduation equation* | | | | | | | |
| $a_{\text{econindex}}$ | Baseline economic index | 0.055 | (0.038) | (0.043) | 0.053 | (0.038) | (0.043) |
| $a_{\text{educ}}$ | Parental schooling (years) | -0.031 | (0.021) | (0.023) | -0.042 | (0.033) | (0.036) |
| $a_{\text{hisp}}$ | Parent Hispanic | -0.077 | (0.156) | (0.170) | -0.070 | (0.154) | (0.169) |
| $a_{\text{page}}$ | Parent age | 0.009 | (0.013) | (0.014) | 0.009 | (0.013) | (0.014) |
| $a_{\text{nkids}}$ | Number of children in household | -0.168 | (0.055) | (0.059) | -0.165 | (0.054) | (0.059) |
| $a_{\text{coh2}}$ | Cohort 2 | 0.392 | (0.178) | (0.186) | 0.389 | (0.179) | (0.189) |
| $a_{\text{coh3}}$ | Cohort 3 | 0.160 | (0.178) | (0.190) | 0.169 | (0.177) | (0.190) |
| $a_{\text{child male}}$ | Male child | -0.069 | (0.147) | (0.156) | -0.040 | (0.151) | (0.162) |
| $a_{\text{chage}}$ | Child age at pretest | -0.229 | (0.243) | (0.262) | -0.279 | (0.249) | (0.270) |
| $a_{K_{i0}\geq\text{med}}$ | $\mathbf{1}\{\text{above-median baseline knowl.}K_{i0}\}$ | – | | | 0.164 | (0.158) | (0.171) |
| $a_{\text{educ}\geq\text{med}}$ | $\mathbf{1}\{\text{above-median parental schooling}\}$ | – | | | 0.089 | (0.224) | (0.241) |
| $\sigma_\omega$ | Std. dev. of graduation shock $\omega_i$ | 0.406 | (0.126) | (0.139) | 0.427 | (0.135) | (0.162) |
| *Panel B: Investment equation* | | | | | | | |
| $\gamma_{Z_X,\text{econ}}$ | Baseline economic index | -0.034 | (0.021) | (0.022) | -0.034 | (0.021) | (0.022) |
| $\gamma_{Z_X,\text{educ}}$ | Parental schooling (years) | 0.015 | (0.008) | (0.008) | 0.015 | (0.008) | (0.008) |
| $\gamma_{Z_X,\text{hisp}}$ | Parent Hispanic | -0.206 | (0.066) | (0.068) | -0.205 | (0.066) | (0.068) |
| $\gamma_{Z_X,\text{page}}$ | Parent age | -0.002 | (0.005) | (0.005) | -0.002 | (0.005) | (0.005) |
| $\gamma_{Z_X,\text{nkids}}$ | Number of children in household | -0.033 | (0.032) | (0.032) | -0.033 | (0.032) | (0.032) |
| $\gamma_{Z_X,\text{coh2}}$ | Cohort 2 | -0.062 | (0.080) | (0.083) | -0.062 | (0.080) | (0.083) |
| $\gamma_{Z_X,\text{coh3}}$ | Cohort 3 | 0.004 | (0.080) | (0.084) | 0.004 | (0.080) | (0.084) |
| $\gamma_{Z_X,\text{male}}$ | Male child | -0.087 | (0.057) | (0.057) | -0.087 | (0.057) | (0.057) |
| $\gamma_{Z_X,\text{chage}}$ | Child age at pretest | -0.028 | (0.105) | (0.104) | -0.028 | (0.105) | (0.104) |
| *Panel C: Knowledge production function* | | | | | | | |
| $\beta_{Z_K,\text{econ}}$ | Baseline economic index | 0.661 | (0.397) | (0.410) | 0.646 | (0.383) | (0.400) |
| $\beta_{Z_K,\text{educ}}$ | Parental schooling (years) | 0.484 | (0.180) | (0.191) | 0.473 | (0.257) | (0.269) |
| $\beta_{Z_K,\text{hisp}}$ | Parent Hispanic | 0.494 | (1.389) | (1.427) | 0.878 | (1.369) | (1.405) |
| $\beta_{Z_K,\text{page}}$ | Parent age | 0.170 | (0.085) | (0.087) | 0.162 | (0.085) | (0.087) |
| $\beta_{Z_K,\text{nkids}}$ | Number of children in household | -1.059 | (0.587) | (0.580) | -1.133 | (0.577) | (0.571) |
| $\beta_{Z_K,\text{coh2}}$ | Cohort 2 | 1.073 | (2.117) | (2.182) | 1.051 | (2.101) | (2.146) |
| $\beta_{Z_K,\text{coh3}}$ | Cohort 3 | 4.311 | (2.509) | (2.568) | 4.264 | (2.500) | (2.566) |
| $\beta_{Z_K,\text{male}}$ | Male child | -1.518 | (1.089) | (1.154) | -1.648 | (1.015) | (1.088) |
| $\beta_{Z_K,\text{chage}}$ | Child age at pretest | 8.584 | (2.071) | (2.093) | 8.742 | (2.073) | (2.081) |
| $\beta_{Z_K,K_{i0}\geq\text{med}}$ | $\mathbf{1}\{\text{above-median baseline knowl.}K_{i0}\}$ | – | | | 3.501 | (2.331) | (2.415) |
| $\beta_{Z_K,\text{educ}\geq\text{med}}$ | $\mathbf{1}\{\text{above-median parental schooling}\}$ | – | | | 3.072 | (2.147) | (2.216) |

*Note:* Remaining structural model parameter estimates, continuing from Table 4. Investment measured by regression (baseline) and Bartlett (endline) factor scores. Structural parameters are computed inverting the maps of the reduced-form coefficients; standard errors are from the delta method, with reduced-form standard errors clustered by cohort and school, and from a school-by-cohort cluster bootstrap of the full estimation procedure (2,000 replications).

Table D2: Structural Model Estimates (Control Group Only)

| Parameter | Description | Estimate | SE (delta) | SE (boot.) |
|---|---|---|---|---|
| *Panel A: Investment equation* | | | | |
| $\widetilde{\gamma}_0$ | Intercept (composite) | 1.359 | (0.735) | (0.761) |
| $\gamma_{X_0}$ | Baseline investment $X_{i0}$ | 0.884 | (0.146) | (0.151) |
| $\gamma_{Z_X,\text{econ}}$ | Baseline economic index | -0.014 | (0.023) | (0.025) |
| $\gamma_{Z_X,\text{educ}}$ | Parental schooling (years) | -0.004 | (0.014) | (0.014) |
| $\gamma_{Z_X,\text{hisp}}$ | Parent Hispanic | -0.292 | (0.095) | (0.097) |
| $\gamma_{Z_X,\text{page}}$ | Parent age | -0.004 | (0.007) | (0.007) |
| $\gamma_{Z_X,\text{nkids}}$ | Number of children in household | -0.053 | (0.047) | (0.048) |
| $\gamma_{Z_X,\text{coh2}}$ | Cohort 2 | -0.117 | (0.102) | (0.107) |
| $\gamma_{Z_X,\text{coh3}}$ | Cohort 3 | -0.015 | (0.117) | (0.123) |
| $\gamma_{Z_X,\text{male}}$ | Male child | 0.049 | (0.090) | (0.093) |
| $\gamma_{Z_X,\text{chage}}$ | Child age at pretest | -0.118 | (0.160) | (0.167) |
| *Panel B: Knowledge production function* | | | | |
| $\beta_0$ | Intercept | -3.766 | (14.614) | (14.845) |
| $\beta_{K_0}$ | Baseline knowledge $K_{i0}$ | 0.375 | (0.046) | (0.046) |
| $\beta_{X_1}$ | Investment $X_{i1}$ | 1.724 | (2.903) | (3.012) |
| $\beta_{Z_K,\text{econ}}$ | Baseline economic index | 0.143 | (0.525) | (0.547) |
| $\beta_{Z_K,\text{educ}}$ | Parental schooling (years) | 0.980 | (0.248) | (0.261) |
| $\beta_{Z_K,\text{hisp}}$ | Parent Hispanic | -1.558 | (2.046) | (2.137) |
| $\beta_{Z_K,\text{page}}$ | Parent age | 0.131 | (0.127) | (0.129) |
| $\beta_{Z_K,\text{nkids}}$ | Number of children in household | -0.651 | (0.762) | (0.778) |
| $\beta_{Z_K,\text{coh2}}$ | Cohort 2 | 1.188 | (2.631) | (2.730) |
| $\beta_{Z_K,\text{coh3}}$ | Cohort 3 | 5.286 | (2.820) | (2.918) |
| $\beta_{Z_K,\text{male}}$ | Male child | -1.009 | (1.921) | (1.965) |
| $\beta_{Z_K,\text{chage}}$ | Child age at pretest | 10.018 | (3.050) | (3.011) |

*Note:* Structural model parameter estimates for the control sample only. Investment measured by regression (baseline) and Bartlett (endline) factor scores. Structural parameters are computed inverting the maps of the reduced-form coefficients; standard errors are from the delta method, with reduced-form standard errors clustered by cohort and school, and from a school-by-cohort cluster bootstrap of the full estimation procedure (2,000 replications).

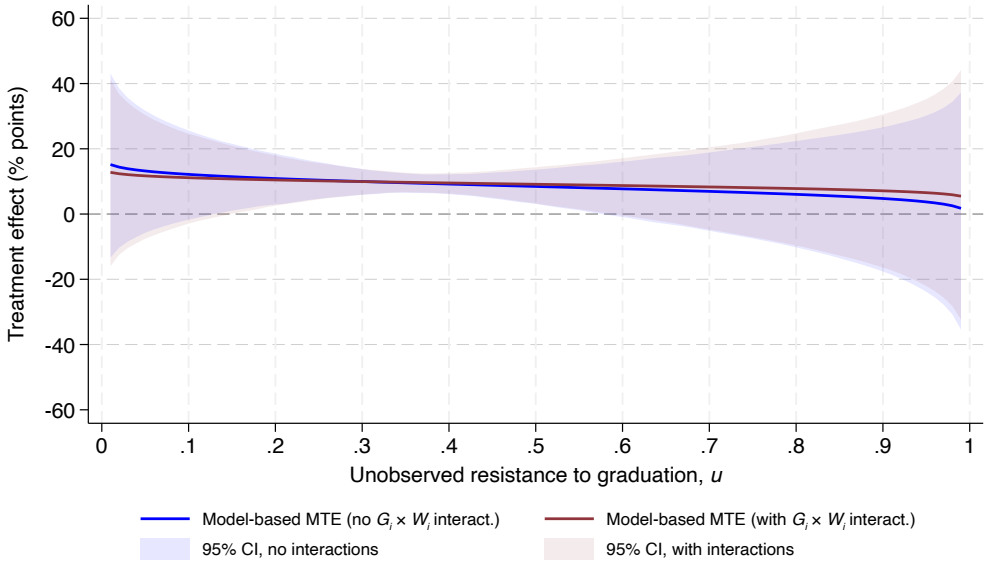
## D.9 Model-Implied MTE

Table D3: Model-Implied MTE Summary

| | (1) No $G_i \times W_i$ interactions | (2) With $G_i \times W_i$ interactions |
|---|---|---|
| $p$-value: flat MTE (bootstrap) | 0.683 | 0.826 |
| $p$-value: flat MTE (analytic, cluster) | 0.681 | 0.830 |
| Compliance rate (graduation rate among offered) | 0.789 | 0.789 |
| Propensity-score support (offered arm) | [0.19, 0.99] | [0.18, 0.98] |
| N (SUR system) | 842 | 842 |
| N (graduation Probit, offered arm) | 454 | 454 |

*Note:* Quantities implied by the estimated model of Section 5. The two $p$-values test that the selection wedge $\rho_1 - \rho_0$ of (25), reported in Table 4, is zero, so that the curve is flat and households do not sort into graduation on unobserved gains, the first from a school-by-cohort cluster bootstrap of the full pipeline and the second analytically with the same clustering. Column (1) restricts the return to graduation to be homogeneous in observables and column (2) lets it vary with $\boldsymbol{W}_i$, as in Table 5, which reports the treatment-effect parameters the model implies. The propensity-score support is the range of the fitted graduation probability in the offered arm. The two sample sizes are those of the seemingly unrelated system and of the graduation probit, which is estimated on the offered arm alone.

Figure D1: Model-Implied Marginal Treatment Effect Curve




*Note:* The solid lines are $\widehat{\mathrm{MTE}}(u)$ from (30), in percentage points. The blue curve is based on the model specification without observed heterogeneity in the return to graduation (Table 4, column 1) while the red curve is based on the model allowing for it (Table 4, column 2), with $\boldsymbol{W}_i$ fixed at their mean levels. The shaded region is the pointwise 95% school×cohort cluster-bootstrap band (70 clusters, 2,000 replications).

# Appendix E Nonparametric Bounds on the Average Treatment Effect

This appendix supports Section 6.1. It sets out the identification problem created by one-sided compliance, states the assumptions under which we bound the average treatment effect, derives the bounds reported in Table 6, and develops the tests of the inequalities that these assumptions imply among observable means.

## E.1 Setup and Identified Moments

We keep the notation of Section 6.1. The offer $T_i$ and graduation $G_i$ are binary, $G_i^*$ is the parent's net gain from graduating in the graduation choice of Section 5.2.2, and one-sided compliance gives $G_i = \mathbf{1}\{T_i = 1\} \cdot \mathbf{1}\{G_i^* > 0\}$. Under the maintained exclusion of the offer, the potential outcomes $K_{i1}(g)$, $g \in \{0,1\}$, are indexed by graduation alone, the observed outcome is $K_{i1} = G_i K_{i1}(1) + (1 - G_i)K_{i1}(0)$, and the score, in percentage points, has logical range $[\underline{K}, \overline{K}]$ = [0, 100].

Write $\pi_t \equiv \Pr(G_i^* > 0 \mid T_i = t)$ for the share of parents who *would* graduate in arm $t$. Assumption RA below makes $T_i$ independent of $G_i^*$, so the two coincide and we write

$$\pi \equiv \Pr(G_i^* > 0) = \Pr(G_i^* > 0 \mid T_i = 1) = \Pr(G_i^* > 0 \mid T_i = 0), \tag{40}$$

whereas one-sided compliance means that this share coincides with the share who *do* graduate only in the treatment arm: $\Pr(G_i = 1 \mid T_i = 1) = \pi$ while $\Pr(G_i = 1 \mid T_i = 0) = 0$. This asymmetry is the source of the identification problem.

Writing $P \equiv \Pr(T_i = 1)$, the data identify exactly three conditional means, which we label by the group they describe:

$$\mu_C \equiv \mathbb{E}[K_{i1} \mid T_i = 0], \qquad \mu_N \equiv \mathbb{E}[K_{i1} \mid G_i = 0, T_i = 1], \qquad \mu_G \equiv \mathbb{E}[K_{i1} \mid G_i = 1, T_i = 1], \tag{41}$$

namely, the control mean, the mean among treated non-graduates, and the mean among treated graduates. It is also convenient to write the treatment-arm mean as

$$\mu_T \equiv \mathbb{E}[K_{i1} \mid T_i = 1] = \pi\mu_G + (1 - \pi)\mu_N,$$

so that the intention-to-treat effect is $\mathrm{ITT} = \mu_T - \mu_C$. In terms of potential outcomes, these conditional means are

$$\begin{aligned}
\mu_G &= \mathbb{E}[K_{i1}(1) \mid G_i^* > 0, T_i = 1], \\
\mu_N &= \mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 1], \\
\mu_C &= \pi\mathbb{E}[K_{i1}(0) \mid G_i^* > 0, T_i = 0] + (1 - \pi)\mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 0].
\end{aligned} \tag{42}$$

The remaining moments are unrestricted by the data, namely the knowledge a never-taker

would have produced had she graduated, $\mathbb{E}[K_{i1}(1) \mid G_i^* < 0, \cdot]$; the knowledge a would-be graduate produces without the program, $\mathbb{E}[K_{i1}(0) \mid G_i^* > 0, \cdot]$, which (42) shows is observed only blended into $\mu_C$; and the whole distribution of $K_{i1}(1)$ in the control arm.

The average treatment effect decomposes by the law of iterated expectations into the effects on those who would and would not graduate,

$$\text{ATE} = \mathbb{E}[K_{i1}(1) - K_{i1}(0)] = \pi \,\text{ATT} + (1-\pi)\,\text{ATU}, \tag{43}$$

with $\text{ATT} = \mathbb{E}[K_{i1}(1) - K_{i1}(0) \mid G_i^* > 0]$ and $\text{ATU} = \mathbb{E}[K_{i1}(1) - K_{i1}(0) \mid G_i^* < 0]$. As we show in Appendix E.3, the experiment identifies the first term but not the second, and hence not the ATE.

## E.2 Assumptions

Beyond the exclusion maintained above, we consider three assumptions and refer to each by name rather than by number. Two of them are the ones our analysis imposes, random assignment (RA) and monotone treatment response (MTR), and every bound reported in Table 6 rests on these two alone. The third, monotone treatment selection (MTS), restricts how families select into graduation. We state it and derive its testable implications alongside the others, but we do not impose it, for the reason given in Section 6.1. Appendix E.5 shows that the tests bearing on it are uninformative in our data, which is why we leave it aside.

**RA. Random assignment.** The offer is independent of the potential outcomes and of the latent willingness to graduate, $\big(K_{i1}(0), K_{i1}(1), G_i^*\big) \perp T_i$. The lottery delivers it, and it gives both $\pi_1 = \pi_0 = \pi$ in (40) and the equality of arm-specific means at a given $G_i^*$. With the exclusion maintained above, it gives

$$\mathbb{E}[K_{i1}(g) \mid G_i^*, T_i = 1] = \mathbb{E}[K_{i1}(g) \mid G_i^*, T_i = 0] = \mathbb{E}[K_{i1}(g) \mid G_i^*], \qquad g = 0,1. \tag{44}$$

Restriction (44) is of the kind Manski and Pepper (2000) call the IV assumption, namely mean independence of the potential outcomes from the instrument. We do not assume it directly. The mean independence in (44) follows from RA for potential outcomes defined under the separately maintained offer exclusion; RA is stronger because independence implies mean independence but not conversely. The exclusion rules out effects of an offer without graduation. Its plausibility depends on how graduation is defined, since parents graduate after attending at least sixteen sessions, so a non-graduate may still have attended several sessions.

**MTR. Monotone treatment response.** Graduating never lowers a child's knowledge:

$$K_{i1}(1) \geq K_{i1}(0) \qquad \text{for all } i.$$

This forces the ATE, the ATT, and the ATU to be non-negative and, in particular, bounds a never-taker's graduate outcome below by her observed non-graduate outcome.

**MTS. Monotone treatment selection.** Parents with a higher latent utility from graduating have weakly better potential outcomes:

$$\mathbb{E}[K_{i1}(g) \mid G_i^* > 0, T_i = t] \geq \mathbb{E}[K_{i1}(g) \mid G_i^* < 0, T_i = t], \qquad g = 0{,}1,\ t = 0{,}1.$$

This is the direction suggested by the observable differences of Section 3.4, where graduates enter with higher baseline scores and more home resources; the assumption itself concerns potential outcomes and cannot be verified.

### E.3 Identification of the Effect on the Treated

Under RA the ATT is point-identified, which is worth recording because the resulting expression reappears as a bound below. Under (44), treated non-graduates and control never-takers have the same expected outcome, $\mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 1] = \mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 0]$, so their contributions cancel in the difference $\mu_T - \mu_C$ and

$$\begin{aligned} & \mathrm{ITT} = \mu_T - \mu_C = \pi\,\mathrm{ATT} \\ \Longrightarrow \qquad & \mathrm{ATT} = \frac{\mu_T - \mu_C}{\pi} = \frac{\mathbb{E}[K_{i1} \mid T_i = 1] - \mathbb{E}[K_{i1} \mid T_i = 0]}{\Pr(G_i = 1 \mid T_i = 1)}. \end{aligned} \tag{45}$$

Equation (45) is the ratio estimator proposed by Bloom (1984) for experiments in which control units cannot receive treatment. Since one-sided compliance sets $\Pr(G_i = 1 \mid T_i = 0) = 0$, the denominator of (45) is also the first stage of the Wald ratio, so the ATT coincides with the local average treatment effect. The ATU, by contrast, is never identified, because $\mathbb{E}[K_{i1}(1) \mid G_i^* < 0]$ is never observed. By (43) the ATE inherits this failure, which is what the bounds below address.

The decomposition (43) also works in reverse. Since the ATT is identified by (45), any bound on the ATE implies a bound on the ATU,

$$\mathrm{ATU} = \frac{\mathrm{ATE} - \pi\,\mathrm{ATT}}{1 - \pi},$$

so an ATE known to lie in [LB,UB] places the ATU in $\big[(\mathrm{LB} - \pi\,\mathrm{ATT})/(1-\pi), (\mathrm{UB} - \pi\,\mathrm{ATT})/(1-\pi)\big]$. Applied to the preferred bounds of Table 6, this is the offer-margin interval for the effect on non-graduates compared with the graduation-margin sets in Section 6.2.

### E.4 Testable Implications of the Identifying Assumptions

None of RA, MTR and MTS is directly verifiable, since each restricts a counterfactual mean. Each combination does, however, imply an inequality among the three observable means in (41),

Table E1: Average Treatment Effect on the Treated Estimates Comparison

| Specification | ITT | Graduation rate ($\widehat{\pi}$) | ATT |
|---|---|---|---|
| **Panel A: Table 3 specifications** | | | |
| Unadjusted difference in means | 7.41 | 0.7823 | 9.47 |
| + Strata fixed effects | 7.70 | 0.7823 | 9.84 |
| + Baseline test score | 7.58 | 0.7823 | 9.69 |
| + Demographic controls | 7.63 | 0.7823 | 9.75 |
| **Panel B: Tables 6 and E3 specifications** | | | |
| Unadjusted conditional means, experiment sample | 7.41 | 0.7823 | 9.47 |
| Adjusted conditional means, experiment sample | 7.08 | 0.7823 | 9.05 |
| Unadjusted conditional means, model sample | | | |
| *Raw graduation rate, treatment arm* | 6.98 | 0.7885 | 8.85 |
| *Probit-fitted, treatment arm* | 6.98 | 0.7884 | 8.85 |
| *Probit-fitted, both arms* | 6.98 | 0.7965 | 8.76 |
| Adjusted conditional means, model sample | | | |
| *Raw graduation rate, treatment arm* | 6.94 | 0.7885 | 8.80 |
| *Probit-fitted, treatment arm* | 6.94 | 0.7884 | 8.80 |
| *Probit-fitted, both arms* | 6.94 | 0.7965 | 8.71 |

*Note:* Intention to treat and average treatment effect on the treated on the endline JumpStart score, in percentage points, with graduation rates as proportions. Each ATT divides the ITT of its row by the graduation rate beside it, following Bloom (1984) and equation (45). Panel A reproduces columns (1) to (4) of Table 3. Panel B reproduces the variants of Tables 6 and E3, whose ITT is built from the same conditional means as the bounds and equals the lower bound under random assignment and monotone treatment response. Its adjusted rows condition on the controls within a treatment by graduation interaction and carry no strata fixed effects, which is why they differ from Panel A. Graduation rates are the raw share of the treatment arm that graduated, except the fitted rates in the model sample, which come from the graduation probit of the structural model and enter only the bounds on the marginal treatment effect (Appendix F.3).

and a reversal of that inequality falsifies the assumptions that generate it. We derive four such implications and report them, along with an exploratory attendance diagnostic, in Table E2. For the four implication tests we test the null that the implied inequality fails against the alternative that it holds, evaluating the $p$-value at the boundary; a small $p$-value is therefore evidence *in favor* of the assumptions, and a large one means the data cannot rule out a violation. Because the implications are necessary but not sufficient, passing a test is corroboration rather than proof.

**RA + MTR: $\mu_T \geq \mu_C$.** Under RA the ITT equals $\pi$ ATT by (45), so this inequality is equivalent to a non-negative ATT, $\mathbb{E}[K_{i1}(1) \mid G_i^* > 0] \geq \mathbb{E}[K_{i1}(0) \mid G_i^* > 0]$, which MTR delivers. A negative $\mu_T - \mu_C$ is therefore evidence against the pair. This is the statistic in the first row of Table E2.

**RA + MTS: $\mu_C \geq \mu_N$.** The control mean averages would-be graduates and never-takers, as in (42), while $\mu_N$ is a pure never-taker mean. RA equates the never-taker component of $\mu_C$ with $\mu_N$, and MTS makes the would-be-graduate component weakly larger, so $\mu_C \geq \mu_N$. A reversal falsifies the pair.

**MTS + MTR: $\mu_G \geq \mu_N$.** The observed gap between graduates and treated non-graduates equals $\mathbb{E}[K_{i1}(1) \mid G_i^* > 0, T_i = 1] - \mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 1]$. MTR makes the first term at least $\mathbb{E}[K_{i1}(0) \mid G_i^* > 0, T_i = 1]$, and MTS makes that at least the second term. Hence $\mu_G \geq \mu_N$, and $\mu_G < \mu_N$ excludes the pair.

**RA + MTS + MTR: $\mu_G \geq \mu_C$.** By RA, $\mu_G = \mathbb{E}[K_{i1}(1) \mid G_i^* > 0, T_i = 0]$, while MTR makes this at least $\mathbb{E}[K_{i1}(0) \mid G_i^* > 0, T_i = 0]$, and MTS makes that at least $\mathbb{E}[K_{i1}(0) \mid G_i^* < 0, T_i = 0]$. Averaging the last two over the control-arm composition in (42) gives $\mu_G \geq \mu_C$. Note that this statistic is also the upper bound of the tightest interval, so the test simultaneously checks that the interval is non-empty.

**Attendance among non-graduates.** The tests above compare outcomes across assignment or graduation groups. As an exploratory diagnostic within the offered non-graduate group, we examine whether endline knowledge varies with the number of sessions attended. Since these families do not graduate, the maintained offer exclusion equates their observed outcome with $K_{i1}(0)$, so

$$\mathrm{Cov}\left(K_{i1}, \text{Attendance}_i \mid G_i = 0, T_i = 1\right) = \mathrm{Cov}\left(K_{i1}(0), \text{Attendance}_i \mid G_i = 0, T_i = 1\right),$$

A positive attendance gradient would be consistent with a positive association between willingness to graduate and knowledge without graduation, if attendance proxies for willingness. Binary MTS compares would-be graduates with never-takers and imposes no restriction on this within-group covariance, so we do not interpret its sign as a test of MTS.

Table E2: Tests of the Identifying Assumptions Behind the Bounds

| | | No control variables | | Regression adjustment | |
|---|---|---|---|---|---|
| Assumption | Null hypothesis ($H_0$) | Mean diff. | $p$-value | Mean diff. | $p$-value |
| RA + MTR | $\mu_T - \mu_C < 0$ | 7.41 | 0.00 | 7.08 | 0.00 |
| RA + MTS | $\mu_C - \mu_N < 0$ | 3.09 | 0.09 | 0.45 | 0.42 |
| MTR + MTS | $\mu_G - \mu_N < 0$ | 13.42 | 0.00 | 9.62 | 0.00 |
| RA + MTS + MTR | $\mu_G - \mu_C < 0$ | 10.33 | 0.00 | 9.17 | 0.00 |
| Attendance diagnostic | $\mathrm{Cov}(\text{Sessions attended}, K_{i1}) = 0$ | 0.00 | 0.78 | 0.00 | 0.96 |

*Note:* Tests of the inequalities that each combination of assumptions implies among the observable means $\mu_C$ (control), $\mu_N$ (treated non-graduates), $\mu_G$ (treated graduates) and $\mu_T$ (treatment arm) defined in (41). Mean differences are in percentage points of the endline JumpStart score; row 5 reports, in the mean-difference columns, the slope of the endline score on sessions attended among treated non-graduates, which has the sign of their covariance, as an exploratory diagnostic, not a test of MTS. The null in rows 1 to 4 is that the implied inequality is violated, with the $p$-value evaluated at the boundary of the null, so a small $p$-value is evidence in favor of the assumptions; the null in row 5 is zero covariance. Appendix E.4 decomposes each mean difference into the identified ATT and the selection gap $\Delta_{\text{sel}}$. RA = random assignment, MTS = monotone treatment selection, MTR = monotone treatment response. The regression-adjusted columns net out the control variables used to estimate the ITT, and rows 1 to 4 come from a single regression of the endline score on assignment, graduation and the controls.

**The anatomy of the statistics.** Under RA the four statistics share a common structure. Define the selection gap

$$\Delta_{\text{sel}} \equiv \mathbb{E}[K_{i1}(0) \mid G_i^* > 0] - \mathbb{E}[K_{i1}(0) \mid G_i^* < 0],$$

the quantity whose sign MTS restricts. Then

$$\mu_T - \mu_C = \pi\,\mathrm{ATT}, \qquad \mu_C - \mu_N = \pi\,\Delta_{\text{sel}}, \qquad \mu_G - \mu_N = \mathrm{ATT} + \Delta_{\text{sel}}, \qquad \mu_G - \mu_C = \mathrm{ATT} + (1-\pi)\,\Delta_{\text{sel}}.$$

Every statistic is a linear combination of two ingredients, the effect on the treated and the selection gap, both identified under RA but estimated with very different precision. The first statistic loads only on the ATT, the second only on $\Delta_{\text{sel}}$, and the last two mix the ingredients with the ATT carrying most of the weight. This anatomy determines what the tests can and cannot show, and Appendix E.5 reads the results through it.

## E.5 Test Results

The tests confirm every implication the treatment effect carries. The treatment–control gap required by RA + MTR is 7.08 percentage points after covariate adjustment (7.41 without) and is positive at the 1 percent level in both specifications, so the implication of the maintained pair passes decisively. The graduate–non-graduate gap required by MTR + MTS is 9.62 points (13.42 unadjusted) and the graduate–control gap required by the triple combination is 9.17 (10.33), both with $p$-values below 0.01. The adjusted differences all come from one regression, so they satisfy the identities of Appendix E.4 exactly, as the unadjusted ones do. The anatomy of Appendix E.4 explains why these rejections come easily, since all three statistics load on the ATT, which is nine to ten points and precisely estimated, so they reject wherever the treatment effect appears, whatever selection does.

The direct statistic for the selection gap is silent. The control–non-graduate contrast equals $\pi\Delta_{\text{sel}}$. Without covariate adjustment it is 3.09 points and only marginally distinguishable from a violation ($p = 0.089$); with adjustment, the like-for-like counterpart of every other adjusted statistic in the paper, it is 0.45 points ($p = 0.422$), an implied selection gap of about half a point with an interval too wide to sign. The exploratory attendance diagnostic likewise shows no detectable association between endline knowledge and sessions attended among treated non-graduates, as the slope of the endline score on sessions attended is zero to two decimals, with $p$-values of 0.78 without covariate adjustment and 0.96 with it.

Read through the anatomy, the pattern is the only one the data could have produced. The tests measure the treatment effect wherever it appears, and they measure the selection gap at a fraction of a point with no precision. This is not a rejection of MTS, since nothing here refutes monotone selection. It is silence about the only quantity MTS restricts, the same limit Section 6.2 documents at the graduation margin. The structural model returns the same verdict by a different route, in that the selection wedge $\rho_1 - \rho_0$ is weakly identified, and its statistical insignificance is a statement about the data's power, not about the absence of sorting. The conclusion for the bounds follows directly. A bound resting on MTS would owe its extra tightness entirely to a restriction on the graduate outcome of families who would decline the program, which the data never reveal, and the one selection contrast the data do identify is too imprecise to sign. Table 6 therefore reports bounds under RA and MTR alone, and the comparison with the structural estimates in Section 6.1 rests on those.

### E.6 Construction of the Bounds

Every bound follows the same recipe. Applying the law of iterated expectations to $\mathbb{E}[K_{i1}(g) \mid T_i = t]$ for $g = 0,1$ and $t = 0,1$,

$$\mathbb{E}[K_{i1}(g) \mid T_i = t] = \pi\mathbb{E}[K_{i1}(g) \mid G_i^* > 0, T_i = t] + (1-\pi)\mathbb{E}[K_{i1}(g) \mid G_i^* < 0, T_i = t],$$

we replace each identified subpopulation mean by its counterpart in (41) and each unidentified mean by the largest or smallest value the maintained assumptions permit. Averaging over $T_i$ and differencing $\mathbb{E}[K_{i1}(1)]$ against $\mathbb{E}[K_{i1}(0)]$ then gives $\text{LB} \leq \text{ATE} \leq \text{UB}$.

**Worst case.** With no assumptions, every unidentified mean is replaced by $\underline{K}$ or $\overline{K}$:

$$\begin{aligned}
\text{LB}_{\text{WC}} &= (1-P)\big(\underline{K} - \mu_C\big) + P\big[\pi\big(\mu_G - \overline{K}\big) + (1-\pi)\big(\underline{K} - \mu_N\big)\big], \\
\text{UB}_{\text{WC}} &= (1-P)\big(\overline{K} - \mu_C\big) + P\big[\pi\big(\mu_G - \underline{K}\big) + (1-\pi)\big(\overline{K} - \mu_N\big)\big].
\end{aligned}$$

The width is $\overline{K} - \underline{K}$, the full range of the score, so these bounds are necessarily uninformative about the sign.

Table E3: Nonparametric Bounds on the Average Treatment Effect (Model Estimation Sample)

| | Unadjusted | | Regression adjusted | |
|---|---|---|---|---|
| | Bounds | 95% CI | Bounds | 95% CI |
| Worst case | [–48.34, 51.66] | [–49.77, 53.08] | [–49.00, 51.00] | [–50.35, 52.35] |
| Random assignment (RA) | [–6.92, 14.22] | [–9.84, 16.76] | [–7.56, 13.59] | [–10.36, 15.78] |
| Monotone treatment response (MTR) | [0.00, 51.66] | [0.00, 53.08] | [0.00, 51.00] | [0.00, 52.35] |
| RA + MTR | [6.98, 14.22] | [4.78, 16.76] | [6.94, 13.59] | [5.10, 15.78] |
| ITT estimate | 6.98 | [4.23, 9.73] | 6.94 | [4.63, 9.24] |
| ATT estimate | 8.85 | [5.36, 12.34] | 8.80 | [5.88, 11.72] |

*Note:* Nonparametric bounds on the average treatment effects and 95% confidence intervals. Numbers estimated exactly as in Table 6, but using the sub-sample used for the model estimation ($N$ = 842).

**Random assignment.** Under RA, $\mathbb{E}[K_{i1}(0)] = \mu_C$ is point-identified and $\mathbb{E}[K_{i1}(1) \mid G_i^* > 0] = \mu_G$, so the only free object is the never-taker's graduate outcome:

$$\mathrm{LB}_{\mathrm{RA}} = \pi\mu_G + (1-\pi)\underline{K} - \mu_C, \qquad \mathrm{UB}_{\mathrm{RA}} = \pi\mu_G + (1-\pi)\overline{K} - \mu_C. \tag{46}$$

The width collapses from $\overline{K} - \underline{K}$ to $(1-\pi)(\overline{K} - \underline{K})$. Equivalently, (46) follows from the decomposition (43) with the ATT identified by (45) and the ATU replaced by its logical range.

**Monotone treatment response.** MTR implies $\mathbb{E}[K_{i1}(0) \mid T_i = 1] \leq \mu_T$, $\mathbb{E}[K_{i1}(1) \mid T_i = 0] \geq \mu_C$, and $\mathbb{E}[K_{i1}(1) \mid T_i = 1] \geq \mu_T$, so the ATE is non-negative and the upper bound stays at its worst-case value:

$$\mathrm{LB}_{\mathrm{MTR}} = 0, \qquad \mathrm{UB}_{\mathrm{MTR}} = \mathrm{UB}_{\mathrm{WC}}.$$

**Combination.** Intersecting the two restrictions gives our preferred bounds, reported in the fourth row of Table 6:

$$\mathrm{LB}_{\mathrm{RA+MTR}} = \mu_T - \mu_C = \mathrm{ITT}, \qquad \mathrm{UB}_{\mathrm{RA+MTR}} = \pi\mu_G + (1-\pi)\overline{K} - \mu_C,$$

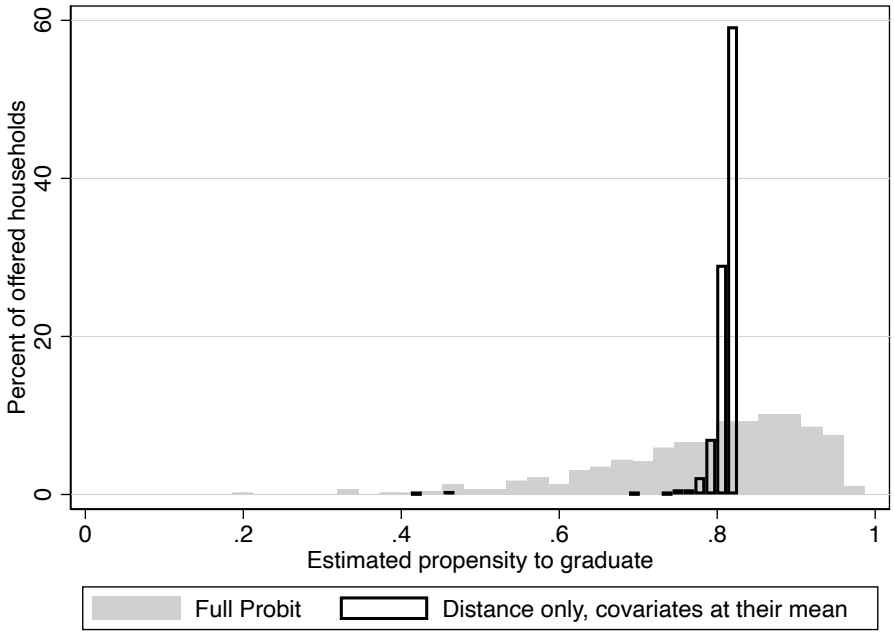
The lower bound deserves emphasis. Under RA the two arms are equated, $\mathbb{E}[K_{i1}(1) \mid T_i = 0] = \mathbb{E}[K_{i1}(1) \mid T_i = 1] = \mathbb{E}[K_{i1}(1)]$, so the MTR restrictions bind through a maximum over arms rather than an average: $\mathbb{E}[K_{i1}(1)] \geq \max\{\mu_C, \mu_T\} = \mu_T$. Equivalently, MTR bounds the never-taker's graduate outcome below by $\mu_N$, so that $\mathbb{E}[K_{i1}(1)] \geq \pi\mu_G + (1-\pi)\mu_N = \mu_T$. Either route gives $\mathrm{LB}_{\mathrm{RA+MTR}} = \mathrm{ITT}$. The upper bound is unchanged from (46), since MTR restricts outcomes only from below; the preferred interval's width is therefore $(1-\pi)(\overline{K} - \mu_N)$, and the only ignorance the maintained assumptions leave is how much a never-taker's child could have gained.

# Appendix F Nonparametric Bounds on the Marginal Treatment Effect

This appendix supports Section 6.2. It quantifies what the distance variation contributes to the graduation propensity and presents the bounds on window averages of the MTE, including the variant that removes the distance moment, that the main text's containment statements refer to.

## F.1 What Distance Contributes

Figure F1: Graduation Propensity: Distance versus the Covariates




*Note:* The shaded histogram is the fitted propensity from the graduation probit of Section 5.4 on the treatment arm only ($N = 454$); the outlined one holds every regressor other than distance at its sample mean and varies distance alone. The two are built from different constructions, so their bin heights are not directly comparable; what is comparable is their dispersion. The fitted propensity has a standard deviation of 0.131 and spans 0.800; moving distance alone gives a standard deviation of 0.034, and even that is inflated by a single household 36.7 kilometers from campus. Appendix Figure F2 shows the same comparison against distance itself.

Distance enters the graduation probit as a natural determinant of participation cost, and Section 5.4 is explicit that identification does not rest on excluding it from the outcome equations. Nevertheless, we document how much work a cost shifter of this kind does. Table F1 and Figures F1 and F2 answer by decomposition. The probit coefficient has the expected sign and is marginally detectable (-0.032 per kilometer, standard error 0.016), but the linear first-stage $F$ on distance is 3.09, far below any conventional threshold for instrument strength. Moving distance alone across its 5th-to-95th percentile range shifts the fitted propensity by 0.035, and by 0.011 across the interquartile range, against a spread of 0.800 when every regressor varies; the probit index has an $R^2$ of 0.008 on distance and 0.464 on baseline investment alone. The raw quartile

Table F1: Distance and Graduation Probability

| *A. First stage: graduation Probit on the offered arm* | |
|---|---|
| Coefficient on home–campus distance (km) | $-0.032$ |
| Cluster-robust standard error (school $\times$ cohort) | $0.016$ |
| $z$-statistic | $-1.99$ |
| Linear first-stage $F$ on distance | $3.09$ |
| *B. Fitted propensity when only distance moves* | |
| 5th percentile of distance ($0.33$ km) | $0.824$ |
| 25th percentile of distance ($0.84$ km) | $0.819$ |
| 50th percentile of distance ($1.42$ km) | $0.815$ |
| 75th percentile of distance ($2.09$ km) | $0.809$ |
| 95th percentile of distance ($4.40$ km) | $0.788$ |
| Maximum ($36.68$ km) | $0.412$ |
| Spread across the interquartile range | $0.011$ |
| Spread across the 5th–95th percentile range | $0.035$ |
| Spread from all Probit variation | $0.800$ |
| $R^2$ of the Probit index on distance | $0.008$ |
| $R^2$ of the Probit index on baseline investment | $0.464$ |

*C. Distance quartiles*

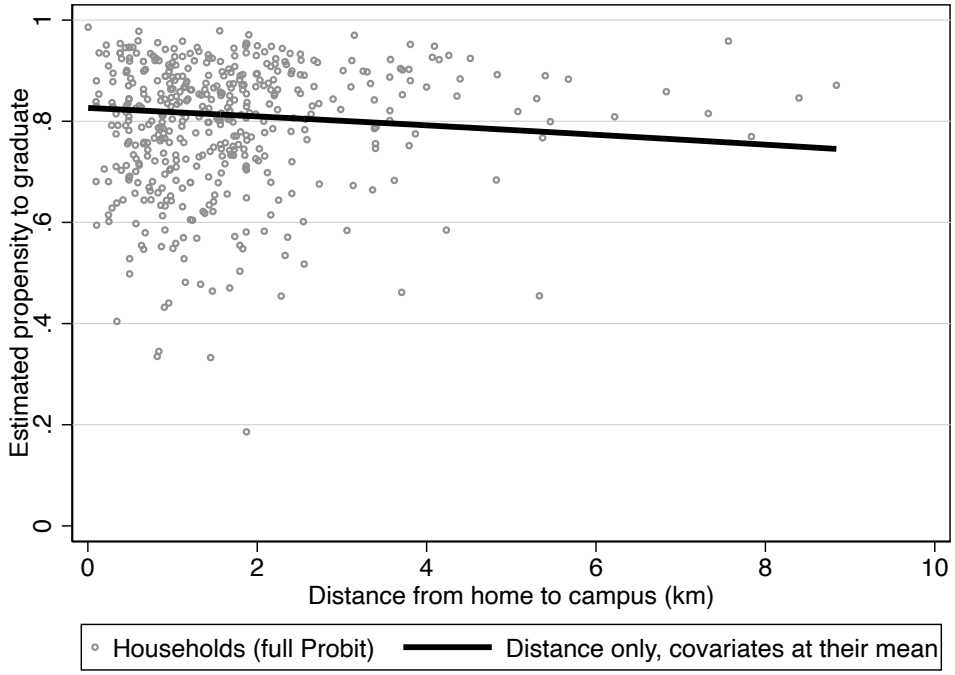
| | $N$ | Mean distance (km) | Graduation rate | Endline score | Fitted propensity |
|---|---|---|---|---|---|
| Q1 | 114 | 0.504 | 0.754 | 0.755 | 0.783 |
| Q2 | 113 | 1.081 | 0.841 | 0.753 | 0.776 |
| Q3 | 115 | 1.726 | 0.748 | 0.737 | 0.786 |
| Q4 | 112 | 4.391 | 0.812 | 0.793 | 0.807 |

*Note:* Offered arm, $N = 454$. Panel A is the coefficient on home-to-campus driving distance in the graduation probit of Section 5.4; distance is the one regressor in that equation that does not also enter the outcome equations. Panel B holds every other regressor at its sample mean and moves distance alone, so it isolates what distance contributes to the spread of the fitted propensity; compare the three spread entries, which give the spread distance generates against the spread the whole probit generates. Panel C is the raw relationship, unadjusted.

means of Panel C are not even monotone. Figure F2 makes the same point visually. The fitted propensities span nearly the unit interval at every distance, so the support of the propensity score is covariate-driven, and the distance profile, covariates held at their means, is nearly flat.

These facts fix how the exhibits of Section 6.2 should be read. The marginal treatment effect curve is not traced by variation that distance induces; it is an implication of the parametric structure, evaluated over support the covariates supply, which is why Section 5.4 reports the selection parameters as weakly identified and why Section 6.2 presents the curve with that caution. In the partial-identification exercise, the moment built on distance is the only one that rests on variation

Figure F2: The Graduation Propensity as a Function of Distance




*Note:* Each circle is a household's fitted propensity from the graduation probit of Section 5.4 on the treatment arm only ($N$ = 454), whose regressors are baseline investment, distance, the household economic index, parental education, ethnicity, age, number of children, cohort dummies, child sex, pretest age, and the missingness flag. The solid line holds every regressor other than distance at its sample mean, so it traces what distance contributes on its own. The vertical scatter at any given distance is covariate-driven variation. Five households beyond ten kilometers are outside the plotted range but are used in every estimate.

excluded from the outcome equations. The tercile moments also bear on the graduation margin, but through variation in the propensity that the covariates generate, which is informative only because the covariates are restricted to shift the level of knowledge and not the gain. Table F2 reports the sets with the distance moment removed. Distance was never asked to serve as a strong instrument; these exhibits record what it was asked to do, and how little of the propensity's variation it accounts for.

## F.2 Bounds on Window Averages of the Marginal Treatment Effect

Figure F3 places the two sets of restrictions side by side, so that what monotone treatment response contributes can be read directly. Without it the lower envelope is negative at every window, so the moments alone never sign the effect window by window. Imposing it lifts the lower envelope to zero or just above, and it also lowers the upper envelope below $u = 0.8$, from 55.5 to 18.6 in the lowest windows and from 32.9 to 10.5 between 0.25 and 0.55. The moments fix weighted integrals of the effect, so a floor on the effect over one range limits how large it can be over another. The upper envelope is unchanged in the window centered at 0.200, where it is 6.7 under

Table F2: Sensitivity of ATE and ATU Bounds to the Distance Moment

| Shape restrictions | ATE | ATU |
|---|---|---|
| *A. Full moment set, including the reduced-form slope on driving distance* | | |
| Outcome bounds only | [3.09, 10.45] | [-19.09, 17.07] |
| *95% CI for the parameter (Imbens–Manski)* | [-3.33, 17.27] | [-47.84, 49.33] |
| Monotone treatment response (MTR) | [8.17, 10.20] | [5.87, 15.84] |
| *95% CI for the parameter (Imbens–Manski)* | [5.74, 13.91] | [2.54, 25.70] |
| *B. Distance moment removed* | | |
| Outcome bounds only | [3.00, 11.45] | [-19.56, 21.96] |
| *95% CI for the parameter (Imbens–Manski)* | [-1.44, 15.30] | [-38.08, 36.24] |
| Monotone treatment response (MTR) | [7.51, 10.76] | [2.62, 18.58] |
| *95% CI for the parameter (Imbens–Manski)* | [5.17, 14.60] | [-1.04, 30.47] |

*Note:* Estimated sets and 95% confidence intervals for the parameter of Imbens and Manski (2004), in percentage points, on the sample, shape restrictions and response functions of Table 7, so that Panel A reproduces that table. Panel A uses the full set of IV-like estimands of Section 6.2, namely the control-arm mean, the mean score of non-graduates and the increment for graduates within each tercile of the fitted propensity among offered families, the unadjusted experimental intention-to-treat effect, and the reduced-form coefficient on home-to-campus driving distance among offered families. Panel B drops the distance moment, the only one built on variation excluded from the outcome equations, and retains the rest. The criterion threshold is held at Panel A's minimum in both panels, so the sets of Panel B contain those of Panel A by construction. The intervals of Panel B rest on 470 bootstrap replications without MTR and 475 with it.

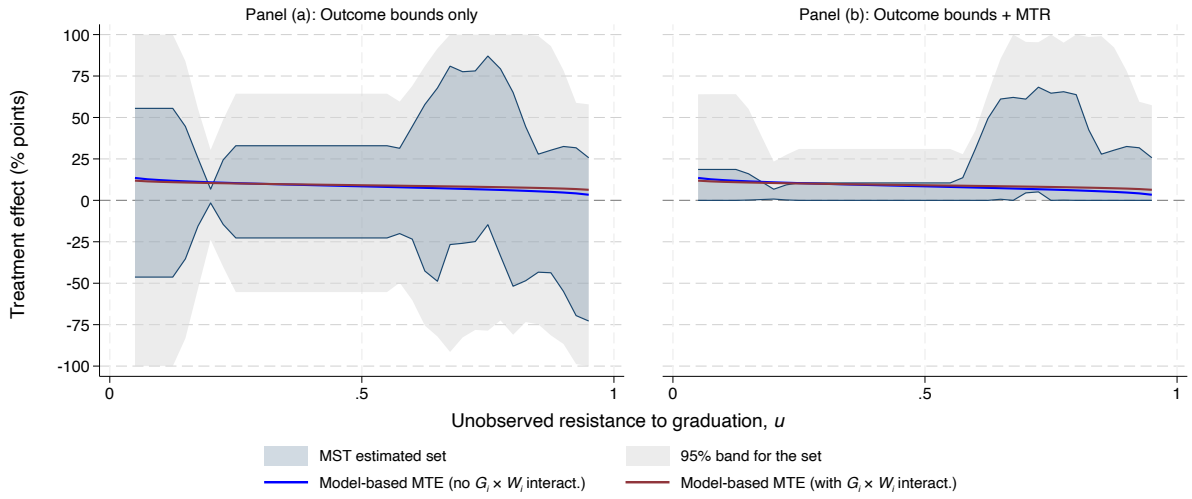
both restrictions, and above 0.85. In the window centered at 0.225 the upper envelope falls from 24.4 to 9.4 once the restriction is imposed, which is why the model curve leaves the set there only under monotone treatment response.

## F.3 Partial Identification of the Graduation Effect

### Method

Section 6.2 sets out the framework. Two details matter for what follows. Covariates enter the response functions, so that $m_g(u,\boldsymbol{c}) = \mathbb{E}[K_{i1}(g) \mid U_{G,i} = u, \boldsymbol{C}_i = \boldsymbol{c}]$, and any moment of the form $\mathbb{E}[s(G_i, T_i, \boldsymbol{C}_i)K_{i1}]$ for a known function $s$, which is what Mogstad et al. (2018) call an IV-like estimand, imposes one linear restriction on the pair $(m_0, m_1)$, so that identification reduces to a linear program. When no pair reproduces the moments exactly, the equality constraints give way to a slack on the minimized criterion, which delivers the least distance set of their equation (27).

Figure F3: Bounds on the MTE Curve, with and without Monotone Treatment Response




*Note:* Bounds following Mogstad et al. (2018) on the MTE averaged over windows of resistance of width 0.10, centered every 0.025 from $u$ = 0.05 to 0.95, computed on the model estimation sample ($N$ = 842). The left panel imposes the outcome range alone and the right panel adds monotone treatment response. The blue and red lines are the model-implied MTE averaged over the same windows, from the estimates in columns (1) and (2) of Table 5, without and with observed heterogeneity in the return to graduation. The shaded gray area is a pointwise 95% percentile band from the school-by-cohort cluster bootstrap, which covers each window's identified set rather than the parameter. Under the outcome range alone it reaches the logical limits of -100 and 100.

### Specification

The estimation sample is the sample used for the structural model, 842 households of which 454 were offered a seat and 388 were not, grouped into 70 school by cohort clusters. The graduation propensity is the probit of the structural model, estimated on the offered arm and fitted for both arms, so that the bounds and the structural estimates rest on the same propensity. We take $m_0$ and $m_1$ to be splines of degree zero on twelve segments of the unit interval, which gives 24 free parameters and imposes no smoothness, and we restrict both to [0, 100], the range of the outcome. Covariates enter with a common coefficient in $m_0$ and $m_1$, so that they shift the level of knowledge without shifting the treatment effect.

The nine IV-like estimands are those listed in Section 6.2, of which eight are linearly independent. The terciles of the fitted propensity within the offered arm are cut at 0.757 and 0.863. The control mean is the one estimand that constrains $m_0$ over the whole unit interval, since a control household cannot graduate whatever its resistance.

These bounds differ from the offer-margin bounds of Section 6.1 in three respects. The first is the information they use. The offer margin uses two identified quantities, the control mean and the effect on the treated, whereas the linear program adds the six cell means within terciles of the fitted propensity and the distance reduced form, which say how outcomes move with the propensity and so bear on the graduation margin directly. The second is that covariates enter both response

functions with the same coefficients, which rules out the covariate-driven heterogeneity in the gain that the offer-margin bounds leave free. The third is the approximation of the response functions. The program searches not over all pairs $(m_0, m_1)$ but over degree-zero splines on twelve segments, and bounds computed over such a family can only be weakly narrower than the nonparametric sharp bounds, so part of the difference in width may reflect that approximation rather than the additional moments. Sets built on different information need not nest, although the parameter lies in their intersection, which here is the narrower of the two.

**Criterion Values and Inference**

The minimized criterion is exactly zero under both sets of restrictions, so the moments are matched exactly and the estimated sets are the sample analogues of the sharp identified sets within the class of response functions searched over, degree-zero splines on twelve segments with covariates entering both functions with common coefficients. Over an unrestricted class the bounds could only be wider. Because the minimum is zero, the slack of equation (27) of Mogstad et al. (2018) plays no role in the estimated sets. It matters in the bootstrap, where the minimum criterion can be positive. Rerunning the optimization on a separate set of 300 cluster bootstrap draws, used only to set the slack, the minimum is strictly positive in 248 of them, or 82.7 percent. In such draws we use the least distance set with an absolute slack, $\kappa_{\text{crit}}$, set to the median of the bootstrap distribution of the minimum, 0.0445.

The intervals in Tables 7 and F2 are the 95 percent confidence intervals for the parameter of Imbens and Manski (2004), built from the bootstrap standard errors of the two bounds over 500 cluster bootstrap replications that resample the 70 school-by-cohort clusters. The bands in Figures 1 and F3 are pointwise percentile bands from the same replications. They cover each window's identified set rather than the parameter and are therefore wider, above all at the upper end, since the bootstrap distribution of the upper bounds is skewed to the right. The unconstrained targets solved in every replication. The targets that hold the effect on the treated fixed, which are those reported, rest on 478 of 500 replications without monotone treatment response and 483 with it, because the search over the multiplier described below returned no solution in the rest. The minimum criterion is strictly positive in 85.0 percent of bootstrap draws without monotone treatment response and in 92.0 percent with it, and the slack can only widen the set in each such draw. The intervals should therefore be interpreted as indicative rather than exact, and the conclusions in the main text rest on the estimated sets.

**Results**

Table 7 reports the identified sets and their confidence intervals for the average effect and for the effect on the never-takers, with the effect on the treated held at 8.762. Here 8.762 is the unadjusted model-sample ITT of 6.979 divided by the probit-fitted graduation probability of 0.7965 averaged over both arms (Appendix Table E1), the denominator that makes the effects on the treated and on

the never-takers partition the population exactly. The choice of denominator moves that quantity by less than a tenth of a percentage point, while the choice of sample moves it by roughly six tenths. The restriction cannot be written as an IV-like moment, since the weight any such moment places on $m_0$ lives on $\{u > p_i\}$, whereas the effect on the treated needs $m_0$ over $\{u \leq p_i\}$. We therefore impose it on the target side, bounding the average effect subject to the effect on the treated equalling 8.762 through the Lagrangian dual of that program, a one-dimensional search over the multiplier in which each evaluation is an ordinary linear program. The restriction binds, narrowing the identified set for the average effect under monotone treatment response from [8.05, 11.98] to [8.17, 10.20]. With the effect on the treated fixed, the bounds on the effect on the never-takers follow from those on the average effect, since $\mathrm{ATE} = \mathrm{ITT} + (1 - \bar{p})\,\mathrm{ATU}$ with $\bar{p}$ = 0.7965. [44] Table F2 reports the same bounds with and without the distance moment. Removing it widens the set for the average effect under monotone treatment response from [8.2, 10.2] to [7.5, 10.8], and its confidence interval from [5.7, 13.9] to [5.2, 14.6]. It widens the set for the effect on the never-takers from [5.9, 15.8] to [2.6, 18.6], and the confidence interval for that effect, [2.5, 25.7] with the distance moment, then includes zero, at [-1.0, 30.5]. The distance moment is therefore what keeps the interval for the effect on the never-takers away from zero, although the estimated set stays positive without it.

The bounds of Figures 1 and F3 are on the MTE averaged over sliding windows of resistance, each of width 0.10 and spaced 0.025 apart, which gives 37 windows, and the model curve is averaged over the same windows before the two are compared. The model lies inside the estimated set in 36 of them without monotone treatment response and in 35 with it, the exceptions being the window centered at $u$ = 0.200 and, once the restriction is imposed, the window at $u$ = 0.225. The pattern is identical for the specifications with and without treatment effect heterogeneity.

[44] Holding the effect on the treated at the value the experiment identifies narrows the estimated set for the never-takers under both sets of restrictions, from [-27.51, 21.57] to [-19.09, 17.07] under the outcome bounds alone and from [5.28, 20.65] to [5.87, 15.84] once monotone treatment response is added. It does not narrow the confidence set under the outcome bounds alone, which widens from [-46.82, 35.85] to [-47.84, 49.33], since the multiplier that enforces the restriction is itself estimated and its sampling variation enters the bound. Both the estimated set and the confidence set narrow once monotone treatment response is imposed, and the confidence set for the average effect narrows under both.

# Appendix G Non-structural Evidence on Mechanisms

This appendix supports the decomposition of Section 6.3, which separates the effect of assignment on endline knowledge into a channel through measured parental investment and a direct channel without imposing the structure of Section 5. Appendix G.1 lets the decomposition differ across groups of families, Appendix G.2 documents its sensitivity to violations of the assumption that baseline investment is a valid instrument for endline investment, and Appendix G.3 derives the bias of that instrument when investment shocks are serially correlated.

## G.1 Group-Specific Direct and Mediated Effects of Assignment

Partition the sample into groups $h = 1, \ldots, H$, let $D_{ih} = \mathbf{1}\{h(i) = h\}$, and consider the system

$$X_{i1} = \sum_h D_{ih} \left(\gamma_{0h} + \gamma_{1h} T_i + \gamma_{2h} X_{i0}\right) + \boldsymbol{\gamma}_3' \boldsymbol{Z}_{1,i} + u_i,$$
$$K_{i1} = \sum_h D_{ih} \left(\beta_{0h} + \beta_{1h} X_{i1} + \beta_{2h} T_i\right) + \boldsymbol{\beta}_3' \boldsymbol{Z}_{2,i} + v_i,$$

where $X_{i1}$ is the endline parental-investment factor score, $X_{i0}$ its baseline counterpart, $K_{i1}$ is endline knowledge, and $T_i$ is the randomized offer. The vectors $\boldsymbol{Z}_{1,i}$ and $\boldsymbol{Z}_{2,i}$ contain the same investment and knowledge controls used in the structural model, with baseline knowledge stacked inside both, so that in our application the two vectors coincide. The parameters $\beta_{1h}$ and $\beta_{2h}$ are, respectively, the return to parental investment and the component of the assignment effect that does not pass through measured investment in group $h$. Full mediation through measured investment arises if $\beta_{2h} = 0$ for all $h$.

Since $T_i$ enters the outcome equation it cannot serve as an instrument for investment; the excluded instruments are the interactions $D_{ih} X_{i0}$, one per endogenous interaction $D_{ih} X_{i1}$, and the system is just identified. The instrument rests on the two conditions of the pooled system of Section 6.3, namely the exclusion that baseline investment affects endline knowledge only through endline investment and uncorrelated measurement errors in $X_{i0}$ and $X_{i1}$, and Appendix G.2 examines the sensitivity of the results to their failure.

We estimate the system by two-stage least squares in one step, with standard errors clustered by school×cohort. Estimating one system rather than $H$ separate regressions delivers the full covariance matrix of $\{\hat{\beta}_{1h}, \hat{\beta}_{2h}\}$, so restrictions across groups can be tested jointly rather than assessed by inspecting the dispersion of group-specific point estimates.

**Choice of partitions.** The number of groups is limited by the strength of the group-specific first stage. Pooled, the coefficient on $X_{i0}$ is 0.802 with a standard error of 0.107, so the first-stage $F$ statistic is approximately 56. Splitting the sample into $H$ groups of roughly equal size divides that statistic by $H$:

| | | | | | | |
|---|---|---|---|---|---|---|
| Groups $H$ | 2 | 3 | 4 | 5 | 9 | 20 |
| Observations per group | 421 | 281 | 211 | 168 | 94 | 42 |
| Implied first-stage $F$ | 27.9 | 18.6 | 14.0 | 11.2 | 6.2 | 2.8 |

We therefore work with median splits on the baseline JumpStart score (our measure of baseline knowledge) and on parent education, the two dimensions of graduation-effect heterogeneity the structural model of Section 5 allows for, as well as median splits on the Bracken baseline score and on the child's age at pretest. Even at two groups the identification is not uniformly strong, so for each group we report Anderson–Rubin confidence sets alongside the Wald standard errors. Because the system is just identified, the Anderson–Rubin sets are valid regardless of first-stage strength, and the comparison between the two indicates how much the conventional intervals can be relied upon.

**Specifications.** Each specification relaxes the pooled system of Section 6.3, the counterpart of the structural specification without graduation–covariate interactions (Table 4, column (1)). We estimate three specifications:

(1) $\beta_{2h}$ unrestricted, with $\gamma_{1h} = \gamma_1$ and $\beta_{1h} = \beta_1$;

(2) $\gamma_{1h}$ and $\beta_{1h}$ unrestricted, with $\beta_{2h} = \beta_2$;

(3) $\gamma_{1h}$, $\beta_{1h}$, and $\beta_{2h}$ all unrestricted.

The three specifications place the burden of the data in different places. Specification (1) frees only the direct effects. Because the freed parameters attach to interactions with the exogenous offer, no additional instrument is required and the first stage remains pooled at its full-sample strength, so the group-specific direct effects are estimated at the precision of intention-to-treat contrasts; this is the specification suited to asking whether the direct effect is present in every group rather than driven by one. Specification (2) frees the endogenous side instead: group-specific first stages and returns require the interacted instruments $D_{ih}X_{i0}$, and with two groups the group-specific first stage falls to roughly half its pooled strength, so precision begins to bind. Specification (3) frees everything; it is the most general form of the across-subgroup comparison, it nests the other two and the pooled system, and it is where the Anderson–Rubin sets matter most.

Specifications (1) and (2) do not nest one another: each relaxes what the other maintains, so together they show which form of homogeneity the data are being asked to carry. Full mediation is the restriction $\beta_{2h} = 0$ in specifications (1) and (3), a joint test across groups, and $\beta_2 = 0$ in the pooled system and in specification (2).

**Results.** Table 8 in the main text reports the pooled decomposition, and Table G1 the over-identification tests. Appendix Tables G2, G3 and G4 report the coefficients, first stages, tests and Anderson–Rubin sets behind each group-specific share, one table per specification.

Table G1: Mediation Analysis: Proportionality Tests Across Subgroups

| | Ratio $\mathrm{ITT}_K/\mathrm{ITT}_X$ | | |
|---|---|---|---|
| | Below median | Above median | $p$-value |
| Baseline JumpStart score | 33.167 | 17.939 | 0.238 |
| Parental schooling (years) | 44.579 | 16.368 | 0.101 |
| Bracken baseline score | 31.390 | 19.608 | 0.307 |
| Child age at pretest (years) | 25.871 | 22.295 | 0.764 |

*Note:* For each splitting variable, the ratio of the intention-to-treat effect of assignment on endline knowledge to its intention-to-treat effect on endline investment, in percentage points per unit of the investment factor, below and above the sample median, and the $p$-value of an over-identification test of their equality (Section 6.3). The test uses only assignment-induced variation. It is a Wald test of the cross-multiplied restriction $\mathrm{ITT}_{K,1}\,\mathrm{ITT}_{X,2} = \mathrm{ITT}_{K,2}\,\mathrm{ITT}_{X,1}$, with standard errors clustered by school×cohort. The knowledge equation carries the controls of equation (32), and the investment equation replaces baseline knowledge with baseline investment. Model estimation sample, $N$ = 842.

**Proportionality test.** The test of Section 6.3 lets the response of investment to assignment differ across groups, $\gamma_{1h}$, and keeps the return $\beta_1$ and the direct effect $\beta_2$ common, so that the ratio of the two intention-to-treat effects in group $h$ is $\beta_1 + \beta_2/\gamma_{1h}$ (footnote 35). Its power comes from differences in $\gamma_{1h}$ across groups, so a failure to reject where the investment response differs little is not evidence of full mediation. We use median splits, $H = 2$, for the reasons given above. Table G1 reports the ratios and the test.

Table G2: IV Mediation Analysis: Group-Specific Direct Effects

| | Mediated share | | Return to investment ($\widehat{\beta}_{1h}$) | | Direct effect ($\widehat{\beta}_{2h}$) | | Effect on investment ($\widehat{\gamma}_{1h}$) | | First-stage $F$ | AR 95% set for $\beta_{1h}$ | $p$-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Share | SE | Coef. | SE | Coef. | SE | Coef. | SE | | | |
| Baseline JumpStart score | | | | | | | | | | | |
| *Below-median* | 0.235 | (0.124) | 6.886 | (2.454) | 6.355*** | (2.017) | 0.284 | (0.053) | 29.4 | [2.169, 12.487] | |
| *Above-median* | 0.341 | (0.190) | 6.886 | (2.454) | 3.784** | (1.554) | 0.284 | (0.053) | 35.8 | [2.169, 12.487] | |
| Full mediation test | | | | | | | | | | | 0.001 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.274 |
| Parental schooling (years) | | | | | | | | | | | |
| *Below-median* | 0.176 | (0.104) | 6.830 | (2.361) | 9.147*** | (2.554) | 0.286 | (0.054) | 11.9 | [2.293, 12.220] | |
| *Above-median* | 0.389 | (0.206) | 6.830 | (2.361) | 3.072** | (1.450) | 0.286 | (0.054) | 77.2 | [2.293, 12.220] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.034 |
| Bracken baseline score | | | | | | | | | | | |
| *Below-median* | 0.213 | (0.109) | 5.906 | (2.401) | 6.215*** | (1.745) | 0.285 | (0.053) | 25.7 | [1.217, 11.310] | |
| *Above-median* | 0.273 | (0.165) | 5.906 | (2.401) | 4.486*** | (1.657) | 0.285 | (0.053) | 36.4 | [1.217, 11.310] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.427 |
| Child age at pretest (years) | | | | | | | | | | | |
| *Below-median* | 0.274 | (0.140) | 6.944 | (2.461) | 5.301*** | (1.729) | 0.288 | (0.054) | 12.6 | [2.215, 12.560] | |
| *Above-median* | 0.303 | (0.153) | 6.944 | (2.461) | 4.601*** | (1.641) | 0.288 | (0.054) | 67.8 | [2.215, 12.560] | |
| Full mediation test | | | | | | | | | | | 0.002 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.717 |

*Note:* $\hat{\beta}_{1h}$ is the return to endline investment, $\hat{\beta}_{2h}$ the direct effect of assignment, and $\hat{\gamma}_{1h}$ the effect of assignment on endline investment, in the cell named at the left. Standard errors are in parentheses, from a school×cohort cluster bootstrap of the whole system (500 replications, medians held fixed) for the mediated share, and clustered by school×cohort for the coefficients. The return and the mediator response are common to both cells here, so the same value is printed twice. First-stage $F$ is the cluster-robust $F$ on $X_{i0}$ in the mediator equation estimated within the cell. The last column is the 95% Anderson–Rubin confidence set for $\beta_{1h}$. Stars on $\hat{\beta}_{2h}$: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Below each block, the joint tests of full mediation and of equal parameters across the two cells. Coefficients on endline knowledge are in percentage points of the score, $\hat{\gamma}_{1h}$ in points of the investment factor.

## G.2 Sensitivity to the Baseline-Investment Exclusion

This subsection documents the sensitivity analysis of Section 6.3, which fixes the violation $\delta$ of equation (33) and estimates the rest. The anchor is the pooled system of Section 6.3.

Table G3: IV Mediation Analysis: Group-Specific Returns to Investment

| | Mediated share | | Return to investment ($\widehat{\beta}_{1h}$) | | Direct effect ($\widehat{\beta}_{2h}$) | | Effect on investment ($\widehat{\gamma}_{1h}$) | | First-stage $F$ | AR 95% set for $\beta_{1h}$ | $p$-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Share | SE | Coef. | SE | Coef. | SE | Coef. | SE | | | |
| Baseline JumpStart score | | | | | | | | | | | |
| *Below-median* | 0.274 | (0.145) | 8.251 | (3.558) | 5.157*** | (1.328) | 0.236 | (0.079) | 29.4 | [-0.361, 19.858] | |
| *Above-median* | 0.254 | (0.127) | 5.288 | (2.250) | 5.157*** | (1.328) | 0.332 | (0.075) | 35.8 | [-1.104, 11.681] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.407 |
| Parental schooling (years) | | | | | | | | | | | |
| *Below-median* | 0.210 | (0.269) | 5.108 | (5.069) | 5.091*** | (1.381) | 0.265 | (0.106) | 11.9 | [-6.093, 34.446] | |
| *Above-median* | 0.314 | (0.135) | 7.603 | (2.392) | 5.091*** | (1.381) | 0.306 | (0.070) | 77.2 | [1.814, 14.902] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.622 |
| Bracken baseline score | | | | | | | | | | | |
| *Below-median* | 0.200 | (0.136) | 5.531 | (3.475) | 5.308*** | (1.288) | 0.241 | (0.070) | 25.7 | [-2.149, 17.600] | |
| *Above-median* | 0.279 | (0.119) | 6.285 | (2.005) | 5.308*** | (1.288) | 0.327 | (0.079) | 36.4 | [1.009, 11.983] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.822 |
| Child age at pretest (years) | | | | | | | | | | | |
| *Below-median* | 0.317 | (0.188) | 8.336 | (4.899) | 4.872*** | (1.475) | 0.271 | (0.083) | 12.6 | [-3.523, 28.443] | |
| *Above-median* | 0.277 | (0.134) | 6.197 | (2.195) | 4.872*** | (1.475) | 0.301 | (0.084) | 67.8 | [0.885, 12.894] | |
| Full mediation test | | | | | | | | | | | 0.001 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.659 |

*Note:* $\hat{\beta}_{1h}$ is the return to endline investment, $\hat{\beta}_{2h}$ the direct effect of assignment, and $\hat{\gamma}_{1h}$ the effect of assignment on endline investment, in the cell named at the left. Standard errors are in parentheses, from a school×cohort cluster bootstrap of the whole system (500 replications, medians held fixed) for the mediated share, and clustered by school×cohort for the coefficients. The direct effect is common to both cells here, so the same value is printed twice. First-stage $F$ is the cluster-robust $F$ on $X_{i0}$ in the mediator equation estimated within the cell. The last column is the 95% Anderson–Rubin confidence set for $\beta_{1h}$. Stars on $\hat{\beta}_{2h}$: $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. Below each block, the joint tests of full mediation and of equal parameters across the two cells. Coefficients on endline knowledge are in percentage points of the score, $\hat{\gamma}_{1h}$ in points of the investment factor.

Table G4: IV Mediation Analysis: Group-Specific Direct Effects and Returns to Investment

| | Mediated share | | Return to investment ($\widehat{\beta}_{1h}$) | | Direct effect ($\widehat{\beta}_{2h}$) | | Effect on investment ($\widehat{\gamma}_{1h}$) | | First-stage $F$ | AR 95% set for $\beta_{1h}$ | $p$-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Share | SE | Coef. | SE | Coef. | SE | Coef. | SE | | | |
| Baseline JumpStart score | | | | | | | | | | | |
| *Below-median* | 0.244 | (0.183) | 8.221 | (3.566) | 6.011*** | (2.190) | 0.236 | (0.079) | 29.4 | [-0.409, 19.852] | |
| *Above-median* | 0.288 | (0.175) | 5.249 | (2.233) | 4.320*** | (1.510) | 0.332 | (0.075) | 35.8 | [-0.626, 11.594] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.404 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.525 |
| Parental schooling (years) | | | | | | | | | | | |
| *Below-median* | 0.113 | (0.263) | 4.667 | (5.058) | 9.746*** | (2.840) | 0.265 | (0.106) | 11.9 | [-6.509, 35.004] | |
| *Above-median* | 0.441 | (0.230) | 7.448 | (2.341) | 2.893* | (1.503) | 0.306 | (0.070) | 77.2 | [1.781, 14.592] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.584 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.039 |
| Bracken baseline score | | | | | | | | | | | |
| *Below-median* | 0.176 | (0.139) | 5.571 | (3.467) | 6.288*** | (1.807) | 0.241 | (0.070) | 25.7 | [-2.092, 17.612] | |
| *Above-median* | 0.322 | (0.170) | 6.314 | (2.016) | 4.340*** | (1.637) | 0.327 | (0.079) | 36.4 | [1.009, 12.043] | |
| Full mediation test | | | | | | | | | | | 0.000 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.824 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.398 |
| Child age at pretest (years) | | | | | | | | | | | |
| *Below-median* | 0.315 | (0.249) | 8.336 | (4.894) | 4.918** | (2.188) | 0.271 | (0.083) | 12.6 | [-3.509, 28.422] | |
| *Above-median* | 0.279 | (0.138) | 6.197 | (2.195) | 4.825*** | (1.487) | 0.301 | (0.084) | 67.8 | [0.883, 12.897] | |
| Full mediation test | | | | | | | | | | | 0.002 |
| Ret. to investment equality test ($\beta_{1h} = \beta_1$) | | | | | | | | | | | 0.659 |
| Direct effect equality test ($\beta_{2h} = \beta_2$) | | | | | | | | | | | 0.968 |

*Note:* $\hat{\beta}_{1h}$ is the return to endline investment, $\hat{\beta}_{2h}$ the direct effect of assignment, and $\hat{\gamma}_{1h}$ the effect of assignment on endline investment, in the cell named at the left. Standard errors are in parentheses, from a school×cohort cluster bootstrap of the whole system (500 replications, medians held fixed) for the mediated share, and clustered by school×cohort for the coefficients. Every parameter is cell-specific here. First-stage $F$ is the cluster-robust $F$ on $X_{i0}$ in the mediator equation estimated within the cell. The last column is the 95% Anderson–Rubin confidence set for $\beta_{1h}$. Stars on $\hat{\beta}_{2h}$: $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. Below each block, the joint tests of full mediation and of equal parameters across the two cells. Coefficients on endline knowledge are in percentage points of the score, $\hat{\gamma}_{1h}$ in points of the investment factor.

**Implementation.** For a fixed $\delta$, define the adjusted outcome

$$K_{i1}^{*}(\delta) \equiv K_{i1} - \delta X_{i0}. \tag{47}$$

Under (33), $X_{i0}$ is a valid instrument for $X_{i1}$ in the equation for $K^*_{i1}(\delta)$ by construction. At each value of $\delta$ we therefore re-estimate the identical two-stage least squares system, $K^*_{i1}(\delta)$ on $T_i$ and $X_{i1}$ with $X_{i1}$ instrumented by $X_{i0}$, with the same controls, sample, and school×cohort clustering as the pooled system, and record the direct effect $\beta_2(\delta)$, the return $\beta_1(\delta)$, and the mediated share $\mathrm{MS}(\delta) = \hat{\gamma}_1\beta_1(\delta)/(\hat{\gamma}_1\beta_1(\delta) + \beta_2(\delta))$; the first stage and the ITT do not involve $\delta$ and are computed once. The grid is symmetric around zero, $\delta \in [-2.2\,\hat{\beta}_1, 2.2\,\hat{\beta}_1]$ in steps of $0.05\,\hat{\beta}_1$.

Confidence intervals along the grid come from the school-by-cohort cluster bootstrap used throughout the paper (500 replications), with $\delta$ held fixed within replications. Because the adjustment (47) is linear, $\beta_2(\delta)$ and $\beta_1(\delta)$ are linear in $\delta$, so the breakdown values reported below have closed forms, which the grid confirms. At $\delta = 0$ the procedure reproduces the pooled system exactly, and the ITT is invariant to $\delta$ by construction. We also express each breakdown value in standard-deviation units, as defined in the note to Table G5.

**The total effect does not depend on $\delta$.** Write the reduced form and the first stage of the two-stage least squares system as $K_{i1} = a_0 + a_T T_i + a_X X_{i0} + \boldsymbol{a}'_Z \boldsymbol{Z}_{2,i} + e_i$ and $X_{i1} = \gamma_0 + \gamma_1 T_i + \gamma_2 X_{i0} + \boldsymbol{\gamma}'_3 \boldsymbol{Z}_{2,i} + u_i$. With one endogenous regressor and one excluded instrument, two-stage least squares coincides with indirect least squares, so that $\hat{\beta}_1 = \hat{a}_X/\hat{\gamma}_2$ and $\hat{\beta}_2 = \hat{a}_T - \hat{\gamma}_1\hat{\beta}_1$, and the total effect $\hat{\beta}_1\hat{\gamma}_1 + \hat{\beta}_2$ equals $\hat{a}_T$. Replacing $K_{i1}$ with $K^*_{i1}(\delta)$ leaves the first stage unchanged, and since $X_{i0}$ is itself a regressor in the reduced form, it changes only the coefficient on $X_{i0}$, from $\hat{a}_X$ to $\hat{a}_X - \delta$. Hence $\hat{\beta}_1(\delta) = (\hat{a}_X - \delta)/\hat{\gamma}_2$ and $\hat{\beta}_2(\delta) = \hat{a}_T - \hat{\gamma}_1(\hat{a}_X - \delta)/\hat{\gamma}_2$, so that each unit of $\delta$ lowers the return by $1/\hat{\gamma}_2$ and raises the direct effect by $\hat{\gamma}_1/\hat{\gamma}_2$, while the total effect stays at $\hat{a}_T$. The mediated share, $\mathrm{MS}(\delta) = \hat{\gamma}_1(\hat{a}_X - \delta)/(\hat{\gamma}_2\hat{a}_T)$, is linear in $\delta$, which is why the breakdown values have closed forms.

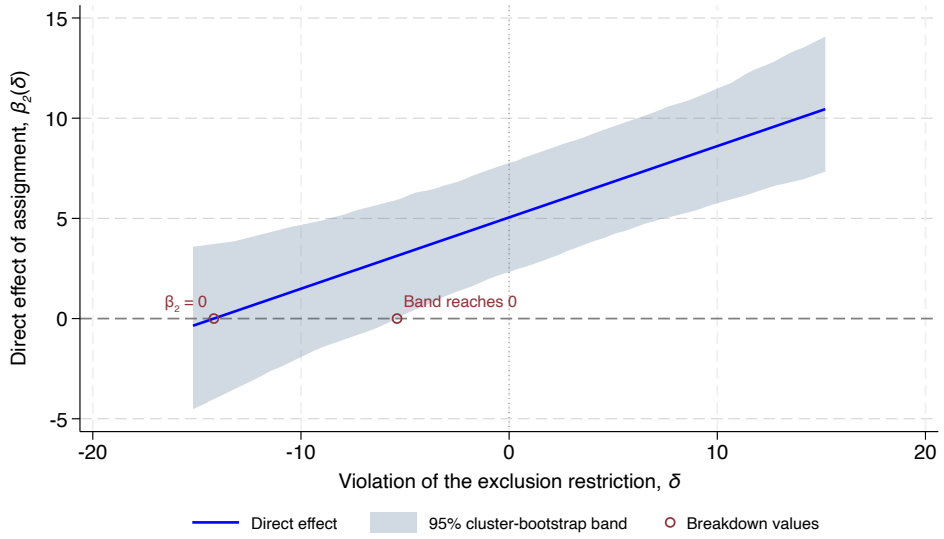
**Results.** Figures G1 and G2 trace the direct effect and the mediated share along the grid, and Table G5 collects the breakdown values. Section 6.3 discusses the direct effect and the point at which the two channels contribute equally. Among the remaining rows, the return to investment loses significance at $\delta$ = 1.6 (0.228 of $\hat{\beta}_1$), and the mediated share falls back to its least-squares value of 0.087 at $\delta$ = 3.8 (0.554 of $\hat{\beta}_1$), so a reinforcing violation of about half of the return removes the whole difference between the least-squares and instrumental-variables decompositions. The share reaches the structural model's share of the effect on the treated, 0.235, at $\delta$ = 0.90, and its share of the average treatment effect, 0.262, at $\delta$ = 0.37, so the two decompositions agree near the exclusion the model maintains, on the same 842 families. The share is significantly above one half only from $\delta$ = -13.0.

Table G5: Breakdown Values for the Baseline-Investment Exclusion

| | $\delta$ | $\delta/\hat{\beta}_1$ | $\delta^{sd}$ | Implied bias in the return |
|---|---|---|---|---|
| Direct effect equals zero (full mediation) | -14.18 | -2.055 | -0.311 | -17.69 |
| its 95% band reaches zero | -5.38 | -0.780 | -0.118 | -6.72 |
| Mediated share equals one half | -4.33 | -0.627 | -0.095 | -5.40 |
| its 95% band reaches one half | -13.02 | -1.887 | -0.286 | -16.25 |
| Share equals the structural model's ATT share (0.235) | 0.90 | 0.130 | 0.020 | 1.12 |
| Share equals the structural model's ATE share (0.262) | 0.37 | 0.054 | 0.008 | 0.47 |
| Share falls back to its OLS value (0.087) | 3.83 | 0.554 | 0.084 | 4.77 |
| Return loses 5% significance | 1.57 | 0.228 | 0.035 | 1.96 |
| Investment channel equals zero | 5.53 | 0.802 | 0.121 | 6.90 |
| $\hat{\beta}_1$ at $\delta = 0$ | | | 6.90 | |
| Direct effect at $\delta = 0$ | | | 5.05 | |
| Mediated share at $\delta = 0$ | | | 0.281 | |

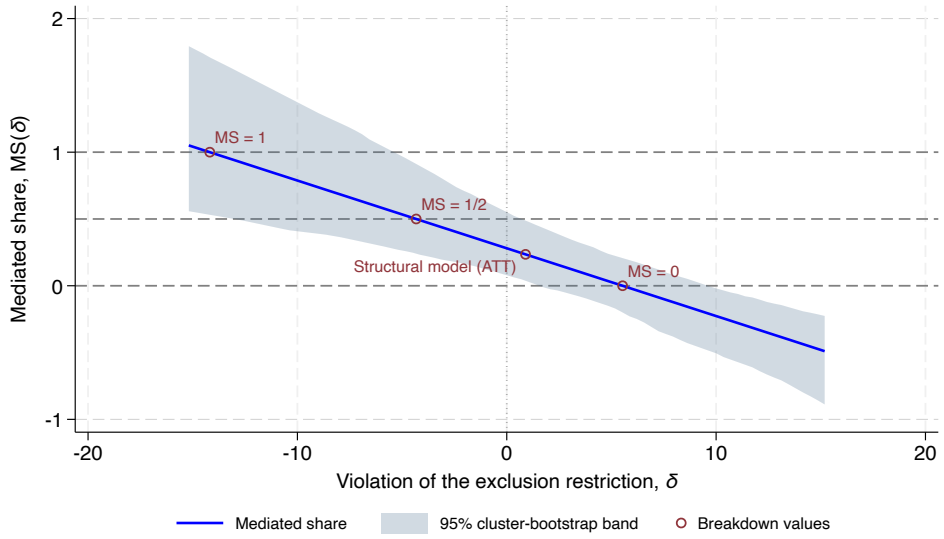
*Note:* Each row gives the value of $\delta$ in equation (33) at which the named event occurs: the direct effect reaching zero and the mediated share reaching one half, each with the robust version at which the 95% band's edge reaches the same value; the share reaching the structural model's shares of the ATT and of the ATE (the indirect channel over each, Table 5, column (1)) and its uncorrected OLS value; and the return losing significance and reaching zero. Four scales: the level of $\delta$ (units of $\beta_1$, percentage points of the endline score per point of the investment factor); the fraction of the estimated return $\hat{\beta}_1$; the standardized violation $\delta^{sd} = \delta\,\sigma_{\tilde{X}_0}/\sigma_v$, with $\sigma_{\tilde{X}_0} = 0.372$ the standard deviation of $X_{i0}$ after partialling the exogenous regressors and $\sigma_v = 16.96$ the structural residual standard deviation at $\delta = 0$; and the implied bias in the return, $\delta/\hat{\gamma}_2$, which by the bound of Appendix G.3 is the floor on the cross-equation slope $|\psi|$ required to generate the violation. The three significance rows are interpolated from the 89-point bootstrap grid of Figures G1 and G2 (500 school×cohort cluster replications, $\delta$ held fixed within replications); the remaining rows are closed forms, which the grid confirms to $10^{-10}$. At $\delta = 0$ the band on the share is [0.079, 0.552]. $N$ = 842.

Figure G1: The Direct Effect of Assignment under Violations of the Exclusion




*Note:* The direct effect $\beta_2(\delta)$ from re-estimating the pooled two-stage least squares system on the adjusted outcome $K^*_{i1}(\delta)$ of equation (47), over the grid $\delta \in [-2.2, 2.2] \times \hat{\beta}_1$ in steps of $0.05\,\hat{\beta}_1$. The shaded region is the pointwise 95% school×cohort cluster-bootstrap band (70 clusters, 500 replications; every replication re-estimates the reduced form and the first stage, and $\delta$ is held fixed). The dashed line marks zero, full mediation; the dotted line marks the maintained $\delta = 0$; the circles mark the breakdown values of Table G5, the point at which the direct effect reaches zero and the point at which the lower edge of its band does. The reduced-form ITT is invariant to $\delta$, so movement along the curve reallocates a fixed total between the two channels.

Figure G2: The Mediated Share under Violations of the Exclusion




*Note:* The mediated share $\mathrm{MS}(\delta) = \hat{\gamma}_1\beta_1(\delta)/(\hat{\gamma}_1\beta_1(\delta) + \beta_2(\delta))$ over the same grid, band and replications as Figure G1. The dashed lines mark zero (no mediation), one half (equal channels) and one (full mediation); the dotted line marks $\delta = 0$, where the share is 0.281. The circles mark the share reaching one, one half, the structural model's share of the ATT, 0.235, and zero (Table G5); it crosses its uncorrected OLS value, 0.087, at $\delta = 3.8$.

### G.3 A Canonical Case: Serially Correlated Investment Shocks

This subsection characterizes the bias of the mediation IV when the shock to the investment policy function is serially correlated and correlated with the shock to the knowledge equation, and shows what that structure implies for the sensitivity parameter $\delta$ of equation (33). To focus on the mechanism, we drop the control vector $\boldsymbol{Z}_{2,i}$, and in the implementation every statement holds conditional on it. We also abstract from measurement error in the factor scores, which is treated separately: this subsection isolates the endogeneity channel.

**Setup.** The system is the simplified counterpart of the pooled system of Section 6.3:

$$X_{i0} = \gamma_{00} + u_{i0}, \tag{48}$$

$$X_{i1} = \gamma_0 + \gamma_1 T_i + \gamma_2 X_{i0} + u_{i1}, \tag{49}$$

$$K_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 T_i + v_{i1}, \tag{50}$$

where (48) and (49) are the investment policy functions at baseline and endline, $\gamma_2$ is structural inertia (the dependence of the endline target on the baseline routine), and (50) is the knowledge equation. The offer $T_i$ does not appear in (48) because it is assigned after baseline. We maintain three assumptions. First, randomization: $T_i$ is independent of $(u_{i0}, u_{i1}, v_{i1})$. Second, AR(1) shocks: the investment shock follows a stationary first-order autoregression,

$$u_{it} = \rho_u \, u_{i,t-1} + \zeta_{it}, \qquad |\rho_u| < 1, \tag{51}$$

with innovations $\zeta_{it}$ that are mean zero, serially uncorrelated, and uncorrelated with $u_{i,t-1}$; stationarity gives $\text{Var}(u_{i0}) = \text{Var}(u_{i1}) = \sigma_u^2$ and $\text{Var}(\zeta_{i1}) = (1 - \rho_u^2)\sigma_u^2$. Third, level correlation across equations: the knowledge shock is correlated with the contemporaneous investment shock through its level,

$$v_{i1} = \psi \, u_{i1} + e_{i1}, \qquad \psi \equiv \frac{\sigma_{uv}}{\sigma_u^2}, \qquad e_{i1} \perp (u_{i0}, u_{i1}, \zeta_{i1}, T_i), \tag{52}$$

so that $\text{Cov}(u_{i1}, v_{i1}) = \sigma_{uv}$; the innovation alternative is treated below. A preliminary fact used throughout:

$$\text{Cov}(u_{i0}, u_{i1}) = \text{Cov}\big(u_{i0}, \rho_u u_{i0} + \zeta_{i1}\big) = \rho_u \, \text{Var}(u_{i0}) + \text{Cov}(u_{i0},\zeta_{i1}) = \rho_u \, \sigma_u^2, \tag{53}$$

where the last equality uses $\text{Cov}(u_{i0},\zeta_{i1}) = 0$.[45]

[45]The example restricts the dynamics of the investment shock but not of the knowledge shock, which may look asymmetric. It is not: $v_{i1}$ enters the system in one period only, so no assumption about its own dynamics is used, and the third assumption above specifies directly the single cross-moment, $\text{Cov}(u_{i0}, v_{i1})$, on which the derivations depend. Serial correlation in the knowledge shock would matter through a different door: in the full system, baseline knowledge is among the controls, and a persistent knowledge shock would make it a contaminated regressor. That is a violation of the value-added specification itself, common to the structural model and the mediation alike rather than particular

**What the first stage identifies.** Substituting (48) into (49) and then (51),

$$X_{i1} = \text{const} + \gamma_1 T_i + \gamma_2 u_{i0} + \underbrace{\rho_u u_{i0} + \zeta_{i1}}_{u_{i1}} = \text{const} + \gamma_1 T_i + (\gamma_2 + \rho_u)\, u_{i0} + \zeta_{i1}. \tag{54}$$

Because $T_i$ is independent of $u_{i0}$, the population regression of $X_{i1}$ on $(1, T_i, X_{i0})$ has coefficient on $X_{i0}$

$$\frac{\mathrm{Cov}(X_{i0}, X_{i1} \mid T_i)}{\mathrm{Var}(X_{i0} \mid T_i)} = \frac{\mathrm{Cov}\big(u_{i0}, \gamma_2 u_{i0} + u_{i1}\big)}{\sigma_u^2} = \frac{\gamma_2 \sigma_u^2 + \rho_u \sigma_u^2}{\sigma_u^2} = \gamma_2 + \rho_u \;\equiv\; \widetilde{\gamma}_2, \tag{55}$$

using (53) for the second term. The first stage therefore identifies only the *sum* of structural inertia and shock persistence. In our data $\widehat{\widetilde{\gamma}}_2 = 0.802$; the split between $\gamma_2$ and $\rho_u$ is not identified with two waves of investment. This is why the analysis of Appendix G.2 is a sensitivity parametrization rather than an estimator.

**The IV estimand.** The mediation IV estimates (50) by two-stage least squares with instruments $(1, T_i, X_{i0})$; the system is just identified. Partialling the exogenous $T_i$, the coefficient on $X_{i1}$ converges to

$$\operatorname{plim} \hat{\beta}_1 = \frac{\mathrm{Cov}(X_{i0}, K_{i1} \mid T_i)}{\mathrm{Cov}(X_{i0}, X_{i1} \mid T_i)}. \tag{56}$$

Substituting (50) into the numerator,

$$\mathrm{Cov}(X_{i0}, K_{i1} \mid T_i) = \beta_1 \,\mathrm{Cov}(X_{i0}, X_{i1} \mid T_i) + \mathrm{Cov}(u_{i0}, v_{i1}), \tag{57}$$

and under (52) and (53),

$$\mathrm{Cov}(u_{i0}, v_{i1}) = \psi \,\mathrm{Cov}(u_{i0}, u_{i1}) = \psi\, \rho_u\, \sigma_u^2 = \rho_u\, \sigma_{uv}. \tag{58}$$

Dividing (57) by $\mathrm{Cov}(X_{i0}, X_{i1} \mid T_i) = (\gamma_2 + \rho_u)\sigma_u^2$ from (55),

$$\operatorname{plim} \hat{\beta}_1 - \beta_1 = \frac{\rho_u\, \sigma_{uv}}{(\gamma_2 + \rho_u)\, \sigma_u^2} = \frac{\rho_u}{\gamma_2 + \rho_u} \cdot \frac{\sigma_{uv}}{\sigma_u^2}. \tag{59}$$

The bias is the product of two factors: the contemporaneous endogeneity slope $\sigma_{uv}/\sigma_u^2$, and the fraction $\rho_u/(\gamma_2 + \rho_u)$ of the instrument's predictive power that comes from the persistent shock rather than from structural inertia. At $\rho_u = 0$ the bias is zero even when $\sigma_{uv} \neq 0$. The direct effect inherits the bias with the opposite sign: the two-stage least squares moment for $T_i$ is $\mathrm{E}\big[T_i(K_{i1} - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 T_i)\big] = 0$; substituting (50) and $\mathrm{E}[T_i v_{i1}] = 0$, and using $\mathrm{E}[T_i X_{i1}]/\mathrm{E}[T_i^2] \to \gamma_1$ after partialling the constant (the $u_{i0}$ term drops because $T_i \perp u_{i0}$), gives

$$\operatorname{plim} \hat{\beta}_2 = \beta_2 + \big(\beta_1 - \operatorname{plim} \hat{\beta}_1\big)\, \gamma_1 = \beta_2 - \frac{\rho_u\, \sigma_{uv}}{(\gamma_2 + \rho_u)\, \sigma_u^2}\, \gamma_1.$$

to the instrument; it would generate biases of the same character and could be examined with adjustments of the same kind.

Reinforcing correlation ($\sigma_{uv} > 0$) with persistence ($\rho_u > 0$) overstates the investment channel and understates the direct effect; compensating correlation does the reverse.

**The innovation case.** Replace (52) with correlation through the innovation only, $v_{i1} = \psi_\zeta \zeta_{i1} + e_{i1}$ with $e_{i1}$ orthogonal to the $u$ process. Then $\text{Cov}(u_{i0}, v_{i1}) = \psi_\zeta \text{Cov}(u_{i0}, \zeta_{i1}) = 0$, so (59) is zero for every $\rho_u$. Purely contemporaneous endogeneity, a shock this period that moves both investment and knowledge, does not contaminate the lagged instrument; only the component of the endogeneity that is itself persistent transmits. The level case and the innovation case bracket the possibilities.

**The mapping to $\delta$.** The sensitivity parameter of (33) is the population projection slope of $v_{i1}$ on the residualized instrument:

$$\delta = \frac{\text{Cov}(X_{i0}, v_{i1} \mid T_i)}{\text{Var}(X_{i0} \mid T_i)} = \frac{\rho_u \, \sigma_{uv}}{\sigma_u^2} = \rho_u \, \psi, \tag{60}$$

using (58). The AR(1) structure therefore gives $\delta$ a decomposition: the persistence of the investment shock times the contemporaneous endogeneity slope. As a consistency check, the IV bias implied by a given $\delta$ follows directly from the violated equation: the numerator of (56) gains the term $\delta \, \text{Var}(X_{i0} \mid T_i) = \delta \sigma_u^2$, so

$$\text{plim}\, \hat{\beta}_1 - \beta_1 = \frac{\delta \, \sigma_u^2}{(\gamma_2 + \rho_u) \, \sigma_u^2} = \frac{\delta}{\widetilde{\gamma}_2}, \tag{61}$$

which coincides with (59) upon substituting (60): the $\delta$-grid of Appendix G.2 and the AR(1) structure are the same object in two parametrizations.

**A bound from the first stage.** Suppose structural inertia is non-negative, $\gamma_2 \geq 0$ (the baseline routine does not structurally reduce the endline target), and shock persistence is non-negative, $\rho_u \geq 0$ (the circumstances that shift investment fade over a year rather than reverse). Then (55) implies $\rho_u \leq \gamma_2 + \rho_u = \widetilde{\gamma}_2$. The sharpest use of this cap runs through the bias itself. Writing $b \equiv \text{plim}\, \hat{\beta}_1 - \beta_1$ and inverting (59), the slope required to produce a given bias is

$$\psi \;=\; \frac{\widetilde{\gamma}_2 \, b}{\rho_u} \qquad \Longrightarrow \qquad |\psi| \;\geq\; |b|, \tag{62}$$

where the inequality substitutes $\rho_u \leq \widetilde{\gamma}_2$. The cross-equation slope must be at least as large, in the same units, as the bias it generates, with equality only in the extreme case $\gamma_2 = 0$; any interior split makes the requirement stricter. Producing a bias equal to the entire estimated return requires an unobserved slope at least as large as the return itself. Persistence high enough to relax the floor carries a price the first stage makes explicit: because $\gamma_2 + \rho_u$ is observed at 0.802, every unit of $\rho_u$ beyond it must be purchased one for one with negative structural inertia.

**Breakdown values in the canonical case.** Each breakdown value $\delta^*$ of Table G5 maps into a bias in the estimated return through (61), $b^* = \delta^*/\hat{\tilde{\gamma}}_2$, and into a floor on the cross-equation slope through (62), $|\psi| \geq |b^*|$. Full mediation sets the direct effect to zero, so that the whole intention-to-treat effect passes through investment and the true return must be $\widehat{\mathrm{ITT}}/\hat{\gamma}_1 = 7.018/0.285 = 24.6$.

**Comparison with ordinary least squares.** For reference, the same structure gives the endogeneity bias of ordinary least squares. Partialling $T_i$, write $\tilde{X}_{i1} = (\gamma_2 + \rho_u)u_{i0} + \zeta_{i1}$ from (54). Under (52), $v_{i1} = \psi(\rho_u u_{i0} + \zeta_{i1}) + e_{i1}$, so

$$\mathrm{Cov}(\tilde{X}_{i1}, v_{i1}) = \psi\Big[(\gamma_2 + \rho_u)\rho_u\,\sigma_u^2 + \mathrm{Var}(\zeta_{i1})\Big] = \psi\,\sigma_u^2\Big[\gamma_2\rho_u + \rho_u^2 + 1 - \rho_u^2\Big] = \psi\,\sigma_u^2\big(1 + \gamma_2\rho_u\big),$$
$$\mathrm{Var}(\tilde{X}_{i1}) = (\gamma_2 + \rho_u)^2\sigma_u^2 + \mathrm{Var}(\zeta_{i1}) = \sigma_u^2\Big[\gamma_2^2 + 2\gamma_2\rho_u + \rho_u^2 + 1 - \rho_u^2\Big] = \sigma_u^2\big(1 + 2\gamma_2\rho_u + \gamma_2^2\big).$$

Hence

$$\operatorname{plim}\hat{\beta}_1^{\mathrm{OLS}} - \beta_1 = \frac{\sigma_{uv}}{\sigma_u^2}\cdot\frac{1 + \gamma_2\rho_u}{1 + 2\gamma_2\rho_u + \gamma_2^2}.$$

Ordinary least squares carries the endogeneity slope essentially at full strength, while the IV bias (59) carries it scaled by $\rho_u/(\gamma_2+\rho_u)$: lagging the instrument eliminates the endogeneity bias exactly when persistence is structural, and inherits it exactly when persistence lives in the shock. This comparison concerns the endogeneity channel only; ordinary least squares additionally suffers the measurement-error attenuation discussed in Section 6.3, which the IV removes regardless of $\rho_u$.