# Intergenerational and Intragenerational Externalities of the Perry Preschool Project*

James J. Heckman
University of Chicago

Ganesh Karapakula
University of Chicago

*James J. Heckman: Center for the Economics of Human Development, 1126 East 59th Street, Chicago, IL 60637; phone: 773-702-0634; fax: 773-702-8490; email: jjh@uchicago.edu. Ganesh Karapakula: Center for the Economics of Human Development, University of Chicago; email: vgk@uchicago.edu. We thank Kurtis Gilliat, John Eric Humphries, Meera Mody, Sidharth Moktan, Tanya Rajan, Azeem Shaikh, Joshua Shea, Winnie van Dijk, and Jin Zhou for providing helpful comments. We also thank Jorge Luis Garcia, Sylvi Kuperman, Juan Pantano, and Anna Ziff for help on related work. We thank Alison Baulos and Lynne Pettler-Heckman for their help in designing the sample survey. We thank Mary Delcamp, Iheoma Iruka, Cheryl Polk, and Lawrence Schweinhart of the HighScope Educational Research Foundation for their assistance in data acquisition, sharing historical documentation, and their longstanding partnership with the Center for the Economics of Human Development. We thank NORC at the University of Chicago for collecting the new data used in this paper. We thank Louise Derman-Sparks and Evelyn K. Moore for discussing and sharing documentation about how the intervention was delivered. This research was supported in part by: the Buffett Early Childhood Fund; NIH Grants R01AG042390, R01AG05334301, and R37HD065072; and the American Bar Foundation. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The web appendix can be found at: http://cehd.uchicago.edu/perry-intergenerational-intragenerational-externalities.

**Abstract**

This paper examines the impact of the iconic Perry Preschool Project on the *children* and *siblings* of the original participants. The children of treated participants have fewer school suspensions, higher levels of education and employment, and lower levels of participation in crime, compared with the children of untreated participants. Impacts are especially pronounced for the children of male participants. These treatment effects are associated with improved childhood home environments. The intergenerational effects arise despite the fact that families of treated subjects live in similar or worse neighborhoods than the control families. We also find substantial positive effects of the Perry program on the siblings of participants who did not directly participate in the program, especially for male siblings.

**Keywords:** externalities; early childhood interventions; spillover effects; intergenerational treatment effects; intragenerational treatment effects

**JEL codes:** C4; I21

James J. Heckman
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: jjh@uchicago.edu

Ganesh Karapakula
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: vgk@uchicago.edu

# 1 Introduction

This paper examines the impact of the Perry Preschool Project — an iconic early childhood program — on the *children* of the original participants. We find significant beneficial intergenerational treatment effects on their education, employment, and crime. This is likely a consequence of the substantial improvement due to intervention in the lives of their parents through late midlife that is documented in a companion paper (Heckman and Karapakula, 2019). Using data from a randomized experiment,[1] we establish that their parents, the original experimental subjects, in the treatment group experienced significant reductions in criminal activity, enhancements in earnings and employment, and better health, executive functioning, and socioemotional skills, compared to those in the control group. Results are particularly strong for male parents.

Our evidence on the outcomes of the original participants is noteworthy because of our fifty-year follow-up. It enables us to measure the outcomes of their adult children. Our research contributes to a limited literature on the intergenerational effects of early interventions.[2]

While we do not conduct a comprehensive mediation analysis, we find suggestive evidence of enhanced environments for the children of original treated participants. They grow up in more stable two-parent families. Their parents have better socioemotional skills, earnings, and employment, and lower participation in crime. However, their childhood neighborhoods are no better and arguably worse than those of the children of control participants. This provides some evidence in support of the importance of family and relative unimportance of zip codes in explaining the observed intergenerational program effects on the children of Perry participants. We also collected data on the siblings of the original participants who did not directly participate in the program. We find evidence of beneficial spillover effects, especially for male siblings.

The plan of this paper is as follows. We first discuss features of the data on the children of

[1] The Perry Preschool Project was a randomized social experiment in the 1960s that aimed to study the lifetime impacts of providing high-quality preschool experiences to socioeconomically and developmentally disadvantaged African-American children. See Heckman and Karapakula (2019) for more details on the intervention.

[2] Barr and Gibbs (2018) and Rossin-Slater and Wüst (2018) study the intergenerational effects of nationwide early childhood programs in the United States and Denmark, respectively, using non-experimental samples. Ours is the first paper to use experimental data to assess second-generation benefits.

participants in Section 2. We propose solutions to address a major limitation of our data—while the original participants were randomized into treatment, their children were not. Section 3 defines the parameters we estimate and why they are useful, in light of data limitations. Section 4 discusses our econometric methods. Section 5 addresses the potential problem of fertility choices by the Perry participants on inference and finds evidence suggesting that it is not particularly important in our sample. Section 6 presents our estimates of intergenerational externalities. Section 7 discusses the early environments of the children of participants. Section 8 discusses the impact of the program on the siblings of participants who were not eligible to participate in the program. Section 9 concludes.

# 2 Data on the Second Generation of Perry Families

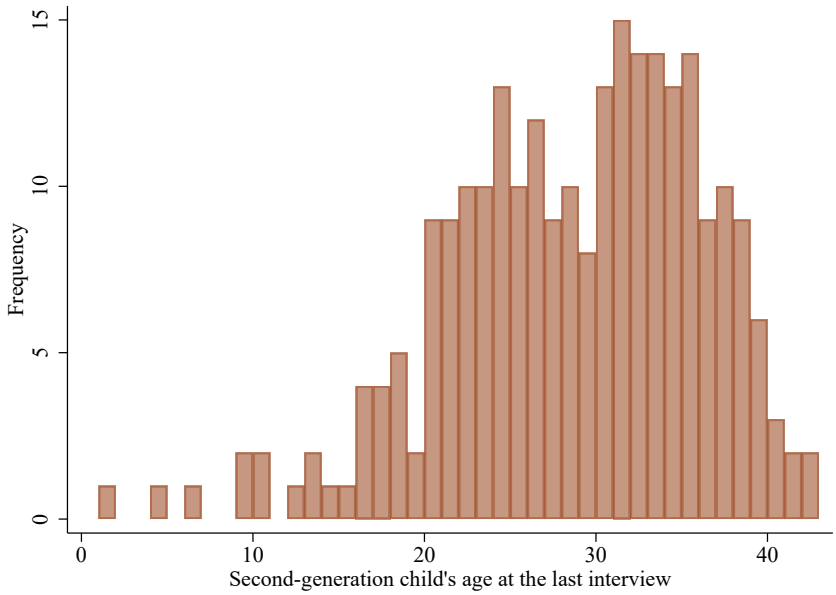
We collected longitudinal data on the the first generation of Perry participants.[3] In their late midlife survey, they were asked questions about their children and siblings. The survey took place over three years: 2014, 2015, and 2016. As a result, we only have data on one cross-section of their children (the second generation) collected in a particular time interval. Figure 1 shows the frequency distribution of the children's ages from the survey. Ages of the children over all interview years range from 1 through 43 years with a mean of about 28 years. Since the survey of the original participants spanned multiple years, information about children of the same birth year was not necessarily collected in the same interview year. As a consequence, the distribution of the second generation's age at interview differs from the density of ages in 2016.[4] The children's ages in 2016 range from 2 to 44 with a mean of about 30 years. By 2016, less than 5% of the children were under age 18 and less than 10% were below the age of 21. In contrast, at the time of the previous follow-up (around 2000), about 63% and 80% were under ages 18 and 21, respectively. As a result, important educational and economic outcomes for most second-generation children were

[3]Heckman and Karapakula (2019) discuss in detail the data collection plan for the late midlife (around age 55) survey of the original participants, who are the parents of the children studied here.

[4]See Figure 1 in the appendix.

unrealized at that time. For this reason, previous Perry studies could not examine adult outcomes of a substantial number of these children in detail.[5]

**Figure 1:** Frequency Distribution of the Second-Generation Children's Ages at the Last Follow-Up



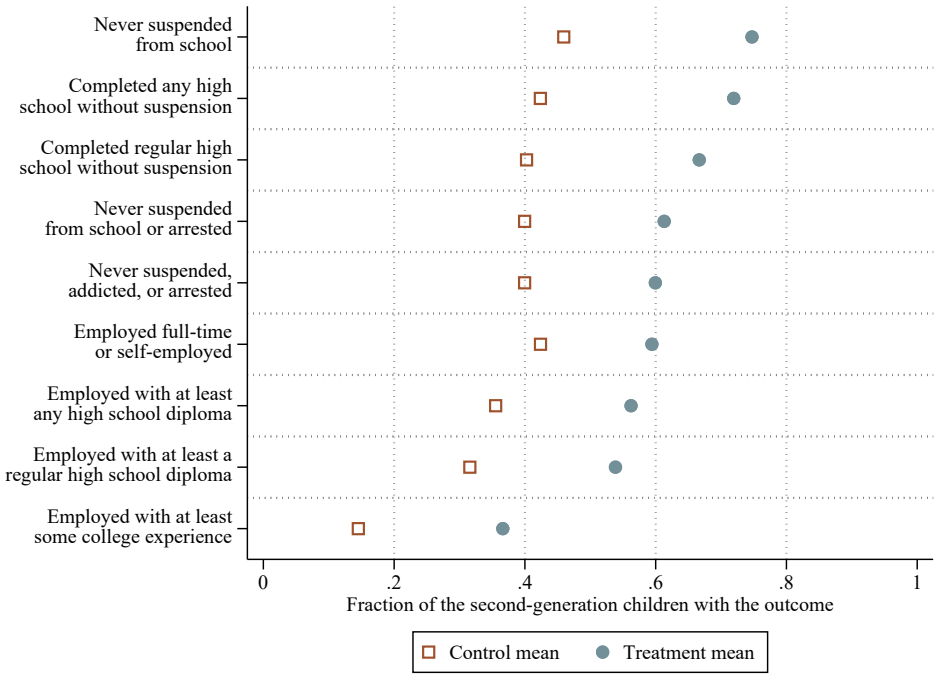

In the late midlife survey, there is information on the adult lives of a majority of the participants' offspring. If the characteristics of participant families differ between treatment and control groups, estimating intergenerational treatment effects solely on the basis of the randomized treatment status of their parents is potentially misleading, due to parental choices about the timing of marriage and fertility. To address this problem, we employ non-experimental methods.

The data on the children of the original participants are not nearly as rich as the data on the original participants.[6] For the second-generation children, we only have categorical or binary data on a small number of outcomes compared to those for the original participants. Nevertheless, the data cover a variety of life domains and give a broad picture of the second generation's outcomes, including school suspensions, arrests, addiction, teenage pregnancy, health, education, and employment.

[5]Only Schweinhart et al. (2005) report the outcomes of the children of participants using a limited sample of second-generation children. The authors state that their data "do not lend themselves to between group comparisons."

[6]We lack administrative crime data, job history data, biomeasures, and so on for the second generation.

**Figure 2:** Statistically Significant Intergenerational Effects at the 10% Worst-Case Level*




**Note*: These estimates of the intergenerational treatment effects are statistically significant at the 10% level using the conservative worst-case test procedures developed in Heckman and Karapakula (2019).

Figure 2 provides a simple summary of the statistically significant intergenerational treatment effects (in terms of mean differences between the experimental groups) for age-based subgroups of the second-generation children.[7] As discussed later, our inferences about these effects survive application of statistically conservative procedures. However, the information presented in Figure 2 alone does not sufficiently account for the problems created by the second-generation sampling procedures. We now document how we address those issues.

[7]The control and treatment means in Figure 2 represent grand means of second-generation children's outcomes, averaged at the participant-level, in the respective experimental groups. Summaries of the first five outcomes and those of the last four outcomes are computed using the subsamples of second-generation children aged 21 and above and those aged 23 and above, respectively. Table 3 contains measures of statistical uncertainty for estimates of the treatment effects. All of the treatment effects in Figure 2 are statistically significant at the 10% level using the worst-case maximum $p$-values, which are conservative measures of statistical uncertainty (Heckman and Karapakula, 2019). The results are robust to a variety of sensitivity checks discussed in the appendix.

# 3 Defining the Parameters of Interest

We define parameters that can feasibly be estimated given the sampling plan. $S_i(b)$ denotes whether participant $i$ had a child in year $b$.[8] $B_i = \{b : S_i(b) = 1\}$ is the set of birth years of the $i$-th participant's children. $C_i(b,a)$ denotes an outcome of interest at age $a$ for the $i$-th participant's child born in year $b$. We define $D_i$ to be the treatment status of the original participant parent.

Define $S_i^d(b)$, $B_i^d$, and $C_i^d(b,a)$, respectively, as the counterfactual outcomes underlying participant $i$'s observed outcomes $S_i(b)$, $B_i$, and $C_i(b,a)$ when $D_i$ is fixed at $d \in \{0,1\}$. The observed outcomes for $i$ are:

$$S_i(b) = (1 - D_i)\, S_i^0(b) + D_i\, S_i^1(b), \tag{1}$$

$$B_i = (1 - D_i)\, B_i^0 + D_i\, B_i^1, \tag{2}$$

and

$$C_i(b,a) = (1 - D_i)\, C_i^0(b,a) + D_i\, C_i^1(b,a). \tag{3}$$

The treatment effect at age $a$ for a child born in year $b$ to original participant $i$ is

$$\tau_i(b,a) = C_i^1(b,a) - C_i^0(b,a). \tag{4}$$

With ideal data, it would be desirable to identify and estimate the parameter

$$\bar{\tau}(b,a) = \mathbb{E}[\tau_i(b,a) \mid S_i(b) = 1], \tag{5}$$

the average treatment effect for the age-$a$ outcome of a child born in year $b$, allowing for het-

[8] We assume that only one child was born in year $b$ and thus abstract away from the case of multiple births in the same year, since this is not an issue in our data. Allowing for multiple births is straightforward.

erogeneity in the treatment effect by age and birth cohort.[9] However, we only observe the child's outcomes in one of the three interview years. Thus, in the best case, we can only identify outcomes for the years and ages we observe: $C_i(I_i - a, a)$, where $I_i \in \{2014, 2015, 2016\}$ is the interview year and $I_i - a = b_i$ for $b_i \in B_i$, the set of birth years actually sampled.[10]

Because of sample limitations, it is necessary to make certain simplifications. For example, suppose that an age-$a$ outcome is constant across birth cohorts so that $C_i(I_i - a, a) = \tilde{C}_i(a)$ across the interview years. It then follows that the parameter $\bar{\tau}(I - a, a)$ can be identified at age $a$ on the common support of the child's age at follow-up available for the interview years $I$.[11] Figures 2 through 24 in the appendix display estimates of this parameter assuming that the fertility decisions of participants are unrelated to treatment. However, due to small cell sizes, the estimates of this parameter are imprecise and sensitive to the choice of the estimator used.

Given these data limitations, we instead estimate an alternative feasible parameter. Suppose that the outcome of interest is teenage pregnancy, which by definition can be measured for all children aged 19 or older. One parameter of interest is the average treatment effect on teenage pregnancy for children whose age at interview is at least 19.[12] Figure 3 shows the pattern of our available data using a Lexis diagram of $I \times b$ cells, where $I$ is the interview year and $b$ is the birth year, with observable teenage pregnancy outcomes. In this diagram, children born in 1998 or later have not had a chance to realize their full potential teenage pregnancy outcomes regardless of the interview year. Those born in 1997 and 1996 have had that opportunity if their interview years

[9] This is a standard parameter defined in the literature (see Heckman, 1979; Lee, 2009). One might also be interested in conditional average treatment effects based on characteristics such as child's gender. We suppress conditioning variables for simplicity of notation, but not in constructing our estimates. There are a variety of other interesting parameters, such as $\tilde{\tau}(a) = \mathbb{E}[\max\{C_i^1(a,b) : b \in B_i^1\} - \max\{C_i^0(a,b) : b \in B_i^0\} \,|\, |B_i^1| \geq 1, |B_i^0| \geq 1]$, the average difference in the best child outcomes between the treatment regimes. Similarly, one could look at the average treatment effect on the worst outcomes by replacing max with min in the previous definition. Estimates of these parameters are explored in Tables 32 through 70 of the appendix.

[10] In practice, the outcomes of all of participant's children are not always recorded for various reasons. For example, the design of the interview limits the number of children surveyed to 5 for those participants with more than five children.

[11] Identification is subject to the assumptions on the relationship between fertility decisions $S_i(b)$ and treatment.

[12] While the age limit in the example is 19, this choice is arbitrary in general, as one may choose to analyze pregnancy up to age 20 or age 18 instead, for instance. The key notion is that it is defined for an age where subjects have completed the process being studied.

are in the sets $\{2016\}$ and $\{2015, 2016\}$, respectively.[13] In terms of Figure 3, the parameter we estimate is the average treatment effect for the second-generation children in the cells with solid circles.

**Figure 3:** Lexis Diagram Showing Cells with Observable Teenage Pregnancy Outcomes

| $I$ \ $b$ | $\cdots$ | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|
| 2014 | $\cdots$ | ● | ● | ○ | ○ | ○ | ○ | $\cdots$ |
| 2015 | $\cdots$ | ● | ● | ● | ○ | ○ | ○ | $\cdots$ |
| 2016 | $\cdots$ | ● | ● | ● | ● | ○ | ○ | $\cdots$ |

*Note*: This Lexis diagram represents the birth year $b$ along the horizontal axis and the interview year $I$ along the vertical axis. A hollow circle indicates that the individuals in the cell do not have a completely realized teenage pregnancy outcome. A solid circle indicates that the individuals in the cell have a realized teenage pregnancy outcome.

Thus in general for a given lower limit $a^*$ on child's age (19 in the case of teenage pregnancy), our parameter of interest is:

$$\bar{\tau}_{a^*} = \mathbb{E}[\bar{\tau}(b, I_i - b) \,|\, I_i - b \geq a^*], \tag{6}$$

where $b$ is the child's birth year, $I_i$ is the interview year, and $I_i - b$ is the age at interview. This parameter is most meaningful for outcomes such as teenage pregnancy which have the property that the event necessarily occurs at or before age $a^*$ and cannot occur afterwards, i.e., $C_i(b, a) = C_i(b, a^*)$ for all $a \geq a^*$. The parameter is also meaningful for outcomes such as graduation from regular high school without suspension by age 21, even though there may be rare cases where the child may have graduated from regular high school after age 21 but before the year of interview. The parameter is less meaningful for outcomes such as crime and employment at ages beyond those surveyed. Nonetheless, $\tau_{a^*}$ is still interpretable as an average effect in the subsample of second-generation children measured at the time of the survey.

[13] Thus, there are two main reasons determining whether the outcome is realized: timing of fertility and childbearing of the original participants, determining whether a second-generation child is too young to have a realized outcome; and the choice of the year made by the interviewer to administer the survey to a first-generation participant. In the next section, we account for these two types of non-response in estimating the intergenerational treatment effects.

# 4 Estimating Treatment Effects

We estimate average intergenerational treatment effects based on the outcome equation:

$$C_i(b,a) = \alpha + \beta D_i + \gamma' Z_i + \delta' W_i(b,a) + \varepsilon_i(b,a), \tag{7}$$

where $D_i$ is the treatment status of the original participant $i$, $Z_i$ is the vector of his or her pre-program covariates,[14] $W_i(b,a)$ is the vector of characteristics of the child born in year $b$ at age $a$, and $\varepsilon_i(b,a) = \nu_i + u_i(b,a)$, where $\nu_i$ represents a fixed effect for original participant $i$, and $u_i(b,a)$ is an error term specific to the birth cohort and age of the child.[15]

The treatment effect parameter in equation (7) is simple to estimate in the relevant subsample of the original participants under the assumption that the participant fertility choices are unaffected by treatment, i.e., $B_i \perp\!\!\!\perp D_i$, and that $\varepsilon_i(b,a) \perp\!\!\!\perp D_i$,[16] if we ignored the covariate vectors $Z_i$ and $W_i(b,a)$.[17] We could simply take the difference in the means of the treatment and control groups for the outcome $\bar{C}_i$, where $\bar{C}_i$ is the average of all non-missing outcomes $C_i(b,a)$ that satisfy our sample inclusion criteria such that $I_i - b \geq a^*$. This would be equivalent to a simple between-effects estimator of equation (7). We could also add back and control for $Z_i$ using ordinary least squares.

Let $R_i$ indicate whether $\bar{C}_i$ can, in principle, be constructed in the available data. The aforementioned estimators necessarily only use the subsample of participants with $R_i = 1$. They do not take into account three distinct sources of non-response that produce $R_i = 0$. The first source is whether or not the original participant was interviewed in the late midlife survey. Let $R_i^1 = 1$

[14] The pre-program covariates include Stanford–Binet IQ, an index of socioeconomic status, sex, and mother's working status at program entry.

[15] Note that in theory it is possible to allow the model parameters to vary by age and birth cohort so that they are all indexed by $(b,a)$.

[16] An interpretation of this assumption is that $\nu_i$, which includes time-invariant residual parental investments common to all children (after accounting for $D_i$ and $Z_i$), and $u_i(b,a)$, which includes residual age- and child-specific parental investments (after accounting for $D_i$, $Z_i$, and $W_i(b,a)$), are orthogonal to the treatment status.

[17] Note that an addition assumption that $\varepsilon_i(b,a) \perp Z_i$ is required for estimators that do not ignore $Z_i$, and a similar assumption is required for estimators that do not ignore $W_i(b,a)$, although ignorability of $Z_i$ and $W_i(b,a)$ is a much stronger assumption.

denote if so; $R_i^1 = 0$ if not. The second source is whether the participant has at least one child in the relevant age group as of 2016 given that $R_i^1 = 1$. Let $R_i^2 = 1$ denote that the second condition is met, and $R_i^2 = 0$ otherwise. The third requirement is that $\bar{C}_i$ is observed for participants with children in the relevant age group. Let $R_i^3 = 1$ if participant $i$ is in this condition, and $R_i^3 = 0$ otherwise. Thus, a person is in our sample or not depending on $R_i = R_i^1 \cdot R_i^2 \cdot R_i^3$. We account for these three sources of sample inclusion using a form of inverse probability weighting.[18]

The estimator we use to construct the main estimates reported in this paper assumes that (i) $\bar{C}_i \perp\!\!\!\perp R_i^1 \,|\, D_i, Z_i$, (ii) $\bar{C}_i \perp\!\!\!\perp R_i^2 \,|\, R_i^1 = 1, D_i, Z_i$, and (iii) $\bar{C}_i \perp\!\!\!\perp R_i^3 \,|\, R_i^2 = 1, R_i^1 = 1, D_i, Z_i$.[19] We use the augmented inverse probability weighting (AIPW) estimator of the intergenerational treatment effect:

$$\hat{\Pi} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \left( \hat{\pi}_i^1 - \hat{\pi}_i^0 \right), \tag{8}$$

where $\mathcal{P}$ is the index set for the sample and

$$\hat{\pi}_i^d = \breve{C}_i^d + \frac{\mathbb{I}\{R_i^3 = R_i^2 = R_i^1 = 1, D_i = d\}}{\hat{\lambda}_{3i}^d \hat{\lambda}_{2i}^d \hat{\lambda}_{1i}^d \hat{\phi}_i^d} \left( \bar{C}_i^d - \breve{C}_i^d \right), \tag{9}$$

$\breve{C}_i^d$ is the gender-specific ordinary least squares-based estimator of $\mathbb{E}[\bar{C}_i \,|\, Z_i, D_i = d, R_i = 1]$ for $d \in \{0, 1\}$,[20] $\hat{\phi}_i^d$ is an estimator of $\Pr(D_i = d \,|\, Z_i)$, the $i$-th participant's propensity of receiving treatment, $\hat{\lambda}_{1i}^d$ is an estimator of $\lambda_{1i}^d = \Pr(R_i^1 = 1 \,|\, D_i = d, Z_i)$, the propensity of being interviewed at the last follow-up after fixing the treatment status, $\hat{\lambda}_{2i}^d$ is an estimator of $\lambda_{2i}^d = \Pr(R_i^2 = 1 \,|\, R_i^1 =$

[18] See, e.g., Huber (2014).

[19] Part (i) of the assumption states that the outcome is independent of the interview status conditional on the treatment status and pre-program covariates. Part (ii) of the assumption states that fertility and childbearing behaviors are independent of the intergenerational outcome conditional on the interview and treatment statuses and pre-program covariates. Part (iii) of the assumption states that any remaining reasons for a missing outcome, including decisions of the interviewer such as timing of the interview and the design of the survey, are conditionally independent of the intergenerational outcome. The estimator also assumes conditional independence of the counterfactual outcomes and the treatment status, i.e., $(\bar{C}_i^{\,1}, \bar{C}_i^{\,0}) \perp\!\!\!\perp D_i \,|\, Z_i$, where $\bar{C}_i^{\,d}$ is the counterfactual version of $\bar{C}_i$ when $D_i$ is fixed at $d \in \{0, 1\}$. This assumption is valid because of the random assignment of the treatment status conditional on pre-program variables. Additionally, for computational ease, we assume that study participants do not engage in sex-selective abortion. This assumption may be controversial for certain cultures, but we believe this assumption holds among the original participants.

[20] Specifically, $\breve{C}_{di} = (Z_i, 1)'(\sum_{k \in \mathcal{G}_{di}} (Z_k, 1)(Z_k, 1)')^{-1}(\sum_{k \in \mathcal{G}_{di}} (Z_k, 1)\bar{C}_k)$, where $G_i$ indicates whether the $i$-th participant is male and $\mathcal{G}_{di} = \{k : D_k = d, R_k = 1, G_k = G_i\}$, for $d \in \{0, 1\}$.

$1, D_i = d, Z_i)$, the propensity of having a child in the relevant age group after fixing the $i$-th participant's treatment and the interview statuses, and $\hat{\lambda}_{3i}^{d}$ is an estimator of $\lambda_{3i}^{d} = \Pr(R_i^3 = 1 \mid R_i^2 = R_i^1 = 1, D_i = d, Z_i)$, the propensity of non-missingness after fixing the previous states, for $d \in \{0,1\}$.[21] $\bar{C}_i^d$ is the counterfactual version of $\bar{C}_i$ for $d \in \{0,1\}$. The AIPW estimator is known to be asymptotically unbiased and has a double robustness property. It is robust to misspecification of either the weighting denominator or the model for counterfactual outcomes.[22] The estimates reported in the text are based on this estimator. In the appendix, we present estimates from other estimators.[23]

We apply the small-sample statistical methods developed in our companion paper (Heckman and Karapakula, 2019) to make the inferences reported in this paper. In that paper we show that conclusions based on conventional inferential procedures can be misleading. Appendix 2 reports twelve different one-sided single $p$-values for each of the estimates of the treatment effect on outcomes we consider. In the text of this paper, we present two of the twelve $p$-values associated with our preferred AIPW estimator. The first is the conventional asymptotic $p$-value $p_{A,A}$.[24] Our companion paper (Heckman and Karapakula, 2019) finds that the asymptotic $p$-values tend to overstate

[21] All of the denominator probabilities for sample inclusion are estimated using a logit specification and the penalized maximum likelihood method of Greenland and Mansournia (2015), which circumvents the issue of separation in small samples. We use a gender-specific logit specification to estimate $\lambda_{2i}^{1}$ and $\lambda_{2i}^{0}$.

[22] See Kang and Schafer (2007), Lunceford and Davidian (2004), and Robins et al. (1994).

[23] Tables 22 through 31 of the appendix contain these alternative estimates. For each subsample of the second-generation children, Tables 22 through 30 of the appendix report results using the two-step estimator (Heckman, 1979), series estimator (Das et al., 2003), lower limit of Lee (2009) bounds, random-effects estimator, and pooled OLS. The two-step and the series estimators do not use exclusion restrictions and are thus limited. The series estimator uses a linear specification for the outcome equation with regressors including the probability of having a child (in the relevant age group) and its square as well as the pre-program covariates. Lee (2009) bounds use OLS-based imputation for missing outcomes for participants with children in the relevant age group. While these bounds are valid under weaker assumptions than the previous estimators, they do not incorporate pre-program variables except for gender. The random-effects and pooled OLS estimators in these tables do not incorporate child- and age-specific regressors. In addition to these estimates, the tables also present heterogeneous treatment effects based on birth order of the children using the between-effects estimator. Table 31 of the appendix presents random-effects and between-effects estimators that include as regressors the following variables: child's gender, child's age at interview and age in 2016, participant's total number of biological children, and the participant's pre-program variables. Additionally, Tables 68 through 70 of the appendix report estimates from between-effects regressions for which the specification of the outcome equation is quadratic in the participant's total number of children, serving as a size-bias correction that adjusts for differences in family sizes.

[24] It is equal to $\Phi(-|\hat{t}|)$, where $\Phi$ is the standard normal distribution function, and $\hat{t}$ is the standardized estimate (based on the custer-robust asymptotic standard error, accounting for dependence between the sibling clusters of the original participants).

the statistical significance of estimates. In that paper, we use the available partial knowledge of the randomization protocol to identify the worst-case (least favorable) $p$-values based on randomization tests. In this paper, we report a version called the "worst-case maximum $p$-value."[25] In the appendix, we also report adjusted $p$-values based on multiple hypothesis tests developed by Holm (1979) and Romano and Wolf (2005).

# 5 Fertility Decisions of the Perry Participants

The endogeneity of fertility decisions of the original participants might confound any estimated treatment effects. If treated parents have fertility much later in life or have fewer children, samples of the second-generation treatment and control children could be fundamentally altered. Since any differences in the fertility patterns between the experimental treatment and control groups can bias simple estimators of the second-generation treatment effects, it is important to examine how the childbearing in the treatment group differs from that in the control group, and to account for these using methods developed in the previous section. In this section, we show that this is not a major

**Table 1:** Treatment Effects on the Completed Fertility of the Participants

| *Variable* | *Sample* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Indicator of any child | Pooled | 0.8095 | 0.7895 | −0.0439 | 0.2710 | 0.3180 |
| Total number of children | Pooled | 2.4237 | 2.5714 | 0.1673 | 0.3143 | 0.5035 |
| Indicator of any child | Male | 0.7436 | 0.7273 | −0.0390 | 0.3550 | 0.3578 |
| Total number of children | Male | 2.0833 | 2.2813 | 0.1900 | 0.3338 | 0.5526 |
| Indicator of any child | Female | 0.9167 | 0.8750 | −0.0509 | 0.3107 | 0.4620 |
| Total number of children | Female | 2.9565 | 2.9583 | 0.1352 | 0.4045 | 0.5076 |

*Note on the columns*: The column labeled *sample* identifies the gender of the Perry participants in the subsample under consideration. *Pooled* refers to the pooled sample of male and female individuals. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups, respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates. The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

concern in analyzing the Perry data. Fertility decisions of the parents of the children we study are barely affected by treatment. The estimates reported in Table 1 reveal that the treatment and control

[25] See Heckman and Karapakula (2019) for more details.

groups do not differ substantially with respect to the total number of children and the probability of having a child over the life cycle.

**Table 2:** Treatment Effects on the Fertility Timing of the Participants

| *Variable* | *Sample* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Age at first birth | Pooled | 21.4800 | 22.6364 | 1.2239 | 0.1318 | 0.2190 |
| Average delivery age | Pooled | 25.6025 | 26.7605 | 1.5100 | 0.0720 | 0.1592 |
| Age at last birth | Pooled | 30.1702 | 30.9302 | 1.1031 | 0.1986 | 0.4160 |
| Age at first birth | Male | 22.4483 | 24.0870 | 1.0502 | 0.2232 | 0.2572 |
| Average delivery age | Male | 27.2064 | 27.7765 | 0.5513 | 0.3476 | 0.4816 |
| Age at last birth | Male | 32.6154 | 31.0909 | −1.2971 | 0.2524 | 0.3185 |
| Age at first birth | Female | 20.1429 | 21.0476 | 1.4692 | 0.2037 | 0.3606 |
| Average delivery age | Female | 23.6167 | 25.6960 | 2.8635 | 0.0254 | 0.0938 |
| Age at last birth | Female | 27.1429 | 30.7619 | 4.4915 | 0.0035 | 0.0387 |

*Note on the columns*: The column labeled *sample* identifies the gender of the Perry participants in the subsample under consideration. *Pooled* refers to the pooled sample of male and female individuals. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups, respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates. The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

The estimates in Table 2 show that the average age at which the participants have the first child does not differ significantly between the experimental groups. There also do not appear to be statistically significant treatment effects on the average age at which participants had children. However, the age at birth of the last child is significantly higher for treated women than for untreated women. This could potentially create imbalances in the composition of the second-generation children of the treated and untreated families within various age ranges. We account for the effect of this imbalance using the weighting methods developed in the previous section. Doing so only has a minor effect on estimates by and large.

# 6 Intergenerational Treatment Effects

Tables 3 through 10 report estimated intergenerational treatment effects of the Perry program by gender of the original participants and that of their children. The estimates in Table 3 show that the children of treatment group members have more than a 30 percentage point higher probability

**Table 3:** Treatment Effects on Outcomes of Pooled Children of the Pooled Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 21$ | 0.4595 | 0.7473 | 0.3232 | 0.0000 | 0.0102 |
| Completed any high school without suspension | $\geq 21$ | 0.4237 | 0.7194 | 0.3348 | 0.0000 | 0.0057 |
| Completed regular high school without suspension | $\geq 21$ | 0.4026 | 0.6667 | 0.3060 | 0.0000 | 0.0085 |
| Never suspended or arrested | $\geq 21$ | 0.3996 | 0.6131 | 0.2015 | 0.0082 | 0.0628 |
| Never suspended, addicted, or arrested | $\geq 21$ | 0.3996 | 0.5995 | 0.1950 | 0.0098 | 0.0661 |
| Employed full-time or self-employed | $\geq 23$ | 0.4239 | 0.5943 | 0.2602 | 0.0009 | 0.0233 |
| Employed with any high school diploma | $\geq 23$ | 0.3553 | 0.5624 | 0.3078 | 0.0005 | 0.0196 |
| Employed with a regular high school diploma | $\geq 23$ | 0.3158 | 0.5386 | 0.3360 | 0.0002 | 0.0118 |
| Employed with some college experience | $\geq 23$ | 0.1453 | 0.3662 | 0.2272 | 0.0001 | 0.0106 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

of never having been suspended from school, compared with their counterparts in the control group. The same is also true for graduation from regular high school or any other type of high school without suspension. This finding is relevant given that African-American students currently represent 38.7% of students suspended from public schools nationwide, even though they comprise only 15.5% of all public school students.[26] Additionally, this finding is relevant because taxpayers benefit when educational attainment of children increases, and because high economic and social costs are associated with high school dropouts.[27] The treatment effect on the probability of never having been arrested or suspended is lower but still sizable, at about 20 percentage points.

We also find large effects (more than 30 percentage points) on the probability of the children of the original participants being employed with regular or any high school diploma. The treatment effect on full-time- or self-employment is about 26 percentage points and that on employment with some college experience is about 23 percentage points. The results on employment taken with the results on suspensions suggest that the children of subjects in the treatment group likely earn more

[26] United States Government Accountability Office (2018) finds that "this pattern of disproportionate discipline persisted regardless of the type of disciplinary action, level of school poverty, or type of public school these students attended."

[27] See Belfield and Levin (2009), Carroll and Erkut (2009), and Rumberger and Losen (2017).

than those in the control group, perhaps due to enhanced cognitive and noncognitive skills,[28] even though we do not have earnings data for the second-generation children.

**Table 4:** Treatment Effects on Outcomes of Male Children of the Pooled Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| In good health (according to parent) | $\geq 18$ | 0.8167 | 0.9383 | 0.1768 | 0.0094 | 0.0799 |
| Graduated from college | $\geq 23$ | 0.0357 | 0.1133 | 0.1150 | 0.0087 | 0.0510 |
| Employed full-time or self-employed | $\geq 23$ | 0.4776 | 0.6667 | 0.2356 | 0.0192 | 0.0874 |
| Employed with any high school diploma | $\geq 23$ | 0.4074 | 0.5900 | 0.2173 | 0.0285 | 0.0926 |
| Employed with a regular high school diploma | $\geq 23$ | 0.3333 | 0.5367 | 0.2703 | 0.0110 | 0.0627 |
| Employed with some college experience | $\geq 23$ | 0.0357 | 0.3367 | 0.3128 | 0.0000 | 0.0130 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic *p*-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum *p*-value based on approximate randomization tests using the studentized AIPW test statistic.

**Table 5:** Treatment Effects on Outcomes of Female Children of the Pooled Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 21$ | 0.5806 | 0.8185 | 0.3059 | 0.0004 | 0.0463 |
| Completed any form of high school | $\geq 21$ | 0.8548 | 0.9655 | 0.1412 | 0.0124 | 0.0992 |
| Graduated from regular high school | $\geq 21$ | 0.7419 | 0.9138 | 0.3013 | 0.0004 | 0.0509 |
| Completed any high school without suspension | $\geq 21$ | 0.5484 | 0.8185 | 0.3616 | 0.0001 | 0.0244 |
| Completed regular high school without suspension | $\geq 21$ | 0.5000 | 0.7649 | 0.3792 | 0.0001 | 0.0211 |
| Employed with any high school diploma | $\geq 23$ | 0.3611 | 0.5000 | 0.3196 | 0.0067 | 0.0980 |
| Employed with a regular high school diploma | $\geq 23$ | 0.2944 | 0.5000 | 0.4051 | 0.0007 | 0.0530 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic *p*-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum *p*-value based on approximate randomization tests using the studentized AIPW test statistic.

Considering only the outcomes of the male children of participants, we find substantial treatment effects on being in good health, graduating from college, and being employed, as shown in Table 4. There are also treatment effects on being employed with some college experience, and on

[28] See Heckman et al. (2018).

being employed at least as a high school graduate, suggesting that the male children in the treated families likely have higher earnings than their control group counterparts.

However, the intergenerational effects reported in Table 3 do not arise solely from the treatment effects for male children. Many of the results regarding school suspensions come mainly from the female children of the original participants. Table 5 shows that female children are much more likely to graduate from regular or any high school without suspension and then to find employment later.

**Table 6:** Treatment Effects on Outcomes of Female Children of the Male Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 21$ | 0.6333 | 0.8974 | 0.3352 | 0.0033 | 0.0970 |
| Completed any form of high school | $\geq 21$ | 0.8667 | 1.0000 | 0.2083 | 0.0065 | 0.0655 |
| Graduated from regular high school | $\geq 21$ | 0.7333 | 0.9643 | 0.4651 | 0.0000 | 0.0365 |
| Completed any high school without suspension | $\geq 21$ | 0.5667 | 0.8974 | 0.4303 | 0.0003 | 0.0510 |
| Completed regular high school without suspension | $\geq 21$ | 0.5000 | 0.8590 | 0.5025 | 0.0001 | 0.0352 |
| Employed full-time or self-employed | $\geq 23$ | 0.4643 | 0.6212 | 0.3764 | 0.0056 | 0.0881 |
| Employed with any high school diploma | $\geq 23$ | 0.3929 | 0.6212 | 0.5255 | 0.0020 | 0.0987 |
| Employed with a regular high school diploma | $\geq 23$ | 0.3214 | 0.6212 | 0.6145 | 0.0003 | 0.0776 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

**Table 7:** Treatment Effects on Outcomes of Male Children of the Male Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never arrested | $\geq 21$ | 0.4167 | 0.6667 | 0.2141 | 0.0320 | 0.0809 |
| Never addicted or arrested | $\geq 21$ | 0.4167 | 0.6667 | 0.2141 | 0.0320 | 0.0809 |
| Graduated from college | $\geq 23$ | 0.0000 | 0.0769 | 0.0813 | 0.0266 | 0.0574 |
| Employed with some college experience | $\geq 23$ | 0.0000 | 0.2564 | 0.2342 | 0.0034 | 0.0523 |
| Employed with a college degree | $\geq 23$ | 0.0000 | 0.0769 | 0.0813 | 0.0266 | 0.0574 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

Table 6 shows that these effects mainly arise for the female children of the male participants. We do not find any effects on the female children of the female participants.

In contrast, there are significant treatment effects on the male children of the male participants. Table 7 shows that for these children, there is about a 21 percentage point difference between the experimental groups in the probability of never having being arrested. About 8% of the second-generation male children of the male participants in the treatment group are college graduates compared to none in the control group. About 26% of those in the treated families are employed with some college experience, while no such children exist in the untreated families.

**Table 8:** Treatment Effects on Outcomes of Pooled Children of the Male Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 21$ | 0.4759 | 0.7898 | 0.3806 | 0.0005 | 0.0310 |
| Completed regular high school without suspension | $\geq 21$ | 0.3933 | 0.7389 | 0.4346 | 0.0000 | 0.0145 |
| Completed any high school without suspension | $\geq 21$ | 0.4246 | 0.7620 | 0.4178 | 0.0001 | 0.0174 |
| Never arrested | $\geq 21$ | 0.6750 | 0.7972 | 0.2308 | 0.0070 | 0.0799 |
| Never suspended or arrested | $\geq 21$ | 0.4333 | 0.6861 | 0.3326 | 0.0025 | 0.0548 |
| Never addicted or arrested | $\geq 21$ | 0.6750 | 0.7972 | 0.2308 | 0.0070 | 0.0799 |
| Never suspended, addicted, or arrested | $\geq 21$ | 0.4333 | 0.6861 | 0.3326 | 0.0025 | 0.0548 |
| Employed full-time or self-employed | $\geq 23$ | 0.4444 | 0.6438 | 0.2995 | 0.0038 | 0.0431 |
| Employed with any high school diploma | $\geq 23$ | 0.3684 | 0.6281 | 0.3812 | 0.0017 | 0.0358 |
| Employed with a regular high school diploma | $\geq 23$ | 0.3246 | 0.5969 | 0.4031 | 0.0013 | 0.0327 |
| Employed with some college experience | $\geq 23$ | 0.1417 | 0.3052 | 0.1688 | 0.0147 | 0.0799 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

Table 8 shows that there are significant treatment effects on the children, male or female, of the male participants. In contrast, Tables 9 and 10 show that all of the effects on the children of the female participants are concentrated on the second-generation male children. We do not find any substantial effects on the female children of the female participants.

**Table 9:** Treatment Effects on Outcomes of Pooled Children of the Female Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 21$ | 0.4439 | 0.7070 | 0.2421 | 0.0041 | 0.0649 |
| Completed any high school without suspension | $\geq 21$ | 0.4228 | 0.6789 | 0.2175 | 0.0088 | 0.0671 |
| Employed with any high school diploma | $\geq 23$ | 0.3421 | 0.5070 | 0.2042 | 0.0355 | 0.0992 |
| Employed with a regular high school diploma | $\geq 23$ | 0.3070 | 0.4895 | 0.2413 | 0.0130 | 0.0513 |
| Employed with some college experience | $\geq 23$ | 0.1491 | 0.4175 | 0.3096 | 0.0003 | 0.0203 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic *p*-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum *p*-value based on approximate randomization tests using the studentized AIPW test statistic.

**Table 10:** Treatment Effects on Outcomes of Male Children of the Female Participants

| *Outcome* | *Age* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Never suspended from school | $\geq 18$ | 0.2440 | 0.5064 | 0.3115 | 0.0199 | 0.0991 |
| Attended college with or without completing | $\geq 23$ | 0.2308 | 0.4514 | 0.3907 | 0.0019 | 0.0637 |
| Employed with some college experience | $\geq 23$ | 0.0769 | 0.4236 | 0.4238 | 0.0013 | 0.0416 |

*Note*: The column labeled *age* indicates the age restrictions defining the subsamples of second-generation children under consideration. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic *p*-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum *p*-value based on approximate randomization tests using the studentized AIPW test statistic.

# 7 Home Environments of the Second Generation

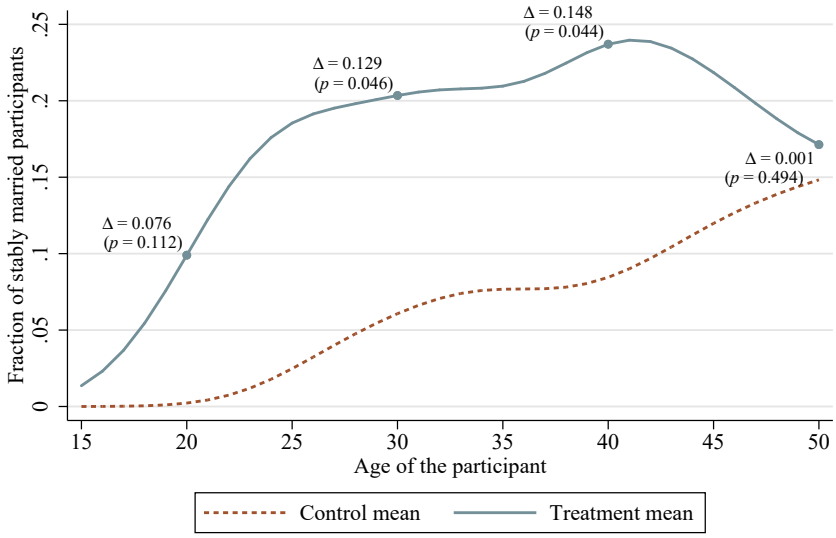
To understand the sources of the substantial intergenerational treatment effects just reported, we examine home environments of the second-generation children. Some recent discussions highlight the role of neighborhoods in shaping children's outcomes (Chetty et al., 2018; Chetty and Hendren, 2018a,b). Another large literature in social science (Almond et al., 2018; Currie and Almond, 2011; Heckman, 2008) points to the role of family structure and family resources in shaping children's outcomes. In this section, we show that the treated Perry participants provide their children more stable two-parent homes. The children of the treated participants excel in various life domains despite growing up in neighborhoods that are similar to or slightly worse off than the neighborhoods

of the control group. The evidence in this section suggests that the home environment matters more than the neighborhood in explaining the intergenerational program effects on the adult outcomes of the children of the Perry participants.

## 7.1 Childhood Family Environments of the Second-Generation Children

Figure 4 shows that the treated Perry men are more likely to be stably married over their life course. We define a stable marriage as one in which the couple was separated for no more than six months regardless of the length of marriage.[29] The higher stable marriage rate of the treated Perry men is likely related to their higher levels of socioemotional skills and executive functioning compared with the untreated men that we document in our companion paper (Heckman and Karapakula, 2019). The stable-marriage rate of the treated men rises through their thirties. It declines slightly

**Figure 4:** Stable Marriage Rate over the Life Course for Male Participants




*Note*: Δ = augmented inverse probability weighting estimate (AIPW) of the treatment effect; $p$ = worst-case maximum $p$-value based on approximate randomization test using studentized AIPW; the control and treatment means are smoothed estimates using the Gaussian kernel with bandwidth of 3.

starting in their early forties. Nevertheless, as indicated in Table 11, stable two-parent childhood environments are 3 times more common until age 18 for the children of the treated participants, and 15 times more common for male children of treated male participants, compared with their control group counterparts. We do not find statistically significant differences in the stable marriage rate

[29] This limited definition is a result of the design of the interview questionnaire.

**Table 11:** Fraction of Time Spent with Stably Married Parents Till Eighteen

| *Children* | *Participants* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Pooled | Pooled | 0.0811 | 0.2459 | 0.1386 | 0.0164 | 0.0578 |
| Male | Pooled | 0.0481 | 0.2512 | 0.1965 | 0.0045 | 0.0618 |
| Female | Pooled | 0.1047 | 0.2466 | 0.1200 | 0.0611 | 0.1748 |
| Pooled | Male | 0.0659 | 0.2341 | 0.1277 | 0.0242 | 0.0952 |
| Male | Male | 0.0174 | 0.2564 | 0.1637 | 0.0057 | 0.0581 |
| Female | Male | 0.0833 | 0.1567 | 0.0553 | 0.2716 | 0.4305 |
| Pooled | Female | 0.0999 | 0.2591 | 0.1540 | 0.1106 | 0.1863 |
| Male | Female | 0.0833 | 0.2449 | 0.2428 | 0.0597 | 0.2218 |
| Female | Female | 0.1311 | 0.3433 | 0.2115 | 0.0573 | 0.1532 |

*Note*: The columns labeled *participants* and *children* identify the gender of the parents (Perry participants) and that of their offspring (the second-generation children) in the subsample under consideration, respectively. *Pooled* refers to the pooled sample of male and female individuals. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

**Table 12:** Average Annual Parental Earnings Till Eighteen

| *Children* | *Participants* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|---|
| Pooled | Pooled | 15.8878 | 28.8892 | 13.4481 | 0.0015 | 0.0338 |
| Male | Pooled | 15.7347 | 31.8143 | 16.6674 | 0.0007 | 0.0406 |
| Female | Pooled | 17.1883 | 23.1280 | 4.1012 | 0.1707 | 0.3201 |
| Pooled | Male | 19.0885 | 39.8030 | 19.7306 | 0.0018 | 0.0371 |
| Male | Male | 21.0772 | 42.3104 | 18.3997 | 0.0142 | 0.1050 |
| Female | Male | 21.5996 | 29.9808 | 6.4368 | 0.1498 | 0.3342 |
| Pooled | Female | 12.2871 | 20.1581 | 4.5787 | 0.1878 | 0.3234 |
| Male | Female | 8.2551 | 22.9329 | 14.2218 | 0.0041 | 0.1665 |
| Female | Female | 12.7769 | 17.1888 | 0.8039 | 0.4418 | 0.4991 |

*Note*: The variable under consideration is average annual earnings (in thousands of 2017 USD) of the parent (the Perry participant) till the child is eighteen years old; this variable is non-missing for a child if the earnings data for the parent are available for at least sixteen of the eighteen years. The columns labeled *participants* and *children* identify the gender of the parents (Perry participants) and that of their offspring (the second-generation children) in the subsample under consideration, respectively. *Pooled* refers to the pooled sample of male and female individuals. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

between the treated and untreated female participants. As a consequence, we are also unable to detect significant treatment effects on the amount of time the children of the female participants

spend with their parents until age 18.

Table 12 shows that the children of the treated male participants have more income resources during their childhoods in terms of parental earnings than their control group counterparts.[30] This is associated with the higher earnings that treated male participants experience in their late twenties and thirties, while engaging less in criminal activity, possibly because of their higher levels of executive functioning and socioemotional skills (Heckman and Karapakula, 2019). These patterns are broadly consistent with the fact that there are more substantial intergenerational treatment effects on the children of the male participants than on the children of the female participants.

## 7.2 Neighborhoods of the Perry Families

We examine the neighborhoods of the Perry families. An influential literature emphasizes the role of neighborhoods in determining intergenerational mobility.[31] We find no significant treatment effects on the choice of the neighborhoods by male participants (around age forty) along several dimensions.[32] These are neighborhoods the children of the participants plausibly grew up in, so the children of both treated and untreated male participants appear to have grown up in similar neighborhoods. Even so, there are substantial intergenerational treatment effects on the children of the male participants, as documented in the previous section.

On the other hand, the children of the treated and untreated female participants on average grow up in different neighborhoods. Compared to the neighborhoods of the control group women, the neighborhoods of the treated women are worse in several dimensions having fewer two-parent homes, lower median household income, lower median family income, more poor families and poor individuals, and higher poverty among families with children. Nonetheless, there are significant beneficial intergenerational treatment effects on the male children of the treated women.

Overall, the children of the treated participants are better off with respect to various life out-

[30] We do not have data on the characteristics of the spouses of the participants, so we are unable to analyze the total household income resources during childhood for the second-generation children.

[31] See Chetty et al. (2018); Chetty and Hendren (2018a,b).

[32] See Appendix 12.

**Table 13:** Treatment Effects on Neighborhood Choices of Female Participants

| *Variable* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|
| Lives in Ypsilanti | 0.5000 | 0.8750 | 0.4325 | 0.0001 | 0.0141 |
| Median age | 32.695 | 30.896 | −1.4487 | 0.0288 | 0.0919 |
| Two parent homes (%) | 0.6162 | 0.5648 | −0.0886 | 0.0053 | 0.0538 |
| Avg. hh. size | 2.5868 | 2.4242 | −0.2252 | 0.0220 | 0.0784 |
| Median hh. income | 44.067 | 40.688 | −6.5739 | 0.0056 | 0.0421 |
| Med. fam. income | 54.445 | 49.351 | −7.9016 | 0.0092 | 0.0876 |
| Median rent | 595.50 | 575.96 | −39.824 | 0.0132 | 0.0672 |
| Med. home value | 1.2298 | 1.0654 | −0.2570 | 0.0108 | 0.0858 |
| Median rooms | 5.0591 | 4.8417 | −0.4069 | 0.0034 | 0.0224 |
| Poor fam. w/ child (%) | 14.100 | 16.638 | 4.1330 | 0.0088 | 0.0811 |

*Note*: *Lives in Ypsilanti* indicates residence in Ypsilanti, Michigan. *Median age* refers to median years of age in the participant's zip code area. *Two parent homes (%)* refers to the percentage of married couples with children among familes with children in the area. *Avg. hh. size* refers to average household size in the participant's zip code area. *Median hh. income* and *med. fam. income* refer to median household income and median family income (in thousands of 1999 dollars) in the area, respectively. *Median rent* refers to median gross rent in the area (in 2000 dollars). *Med. home value* refers to median value (in hundreds of thousands of 2000 dollars) of owner-occupied units in the area. *Median rooms* refers to median number of rooms in the housing units within the participant's zip code area. *Poor fam. w/ child (%)* refers to the percentage of families with children under 18 years in poverty in the area. The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the second-generation children (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response, imbalance in participants' preprogram covariates between the experimental groups, and participants' fertility differences in having a child in the specified age group). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

comes in spite of growing up in neighborhoods that are no better than neighborhoods of the control group members. The evidence in this section combined with that in the previous section suggests that conducive childhood home environments may have reversed any adverse neighborhood effects experienced by the children of the treated participants.

# 8 Effects on Siblings of the Participants

A byproduct of the survey at late midlife is information asked of the original participants on the outcomes of their siblings. We examine the effects of treatment on the siblings of the original participants who did not themselves participate in the experiment.[33] (Siblings eligible for participation in the program when the program was conducted were enrolled as participants.) We do

[33]The literature on intragenerational externalities of early interventions is limited. See, e.g., Ravindran (2018) for a documentation of negative spillover effects on the siblings of participants of an early childhood program in India using a non-experimental sample.

not find statistically significant effects on the siblings of the participants who were born after their families were enrolled in the experiment. However, we find treatment effects for the siblings of the participants who were present at baseline but not eligible to directly participate in the experiment. Table 14 shows that such pre-program siblings, who were present before their families joined the experiment, belonging to the treatment group are more likely to graduate from regular high school and be employed than the pre-program siblings of the untreated participants.

**Table 14:** Treatment Effects on Outcomes of Pooled Pre-Program Siblings of Pooled Participants

| *Outcome* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|
| Graduated from regular high school | 0.7845 | 0.8441 | 0.1309 | 0.0155 | 0.0820 |
| Employed full-time or self-employed | 0.7286 | 0.9147 | 0.1999 | 0.0077 | 0.0482 |
| Employed with a regular high school diploma | 0.5828 | 0.7532 | 0.2147 | 0.0068 | 0.0416 |
| Employed with any high school diploma | 0.6536 | 0.7796 | 0.1876 | 0.0202 | 0.0547 |

*Note*: The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the pre-program siblings (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response and imbalance in participants' preprogram covariates between the experimental groups). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

**Table 15:** Treatment Effects on Outcomes of Male Pre-Program Siblings of Pooled Participants

| *Outcome* | *Untreated mean* | *Treated mean* | *AIPW estimate* | *Asymptotic p-value* | *Worst-case max. p* |
|---|---|---|---|---|---|
| Graduated from regular high school | 0.6932 | 0.8764 | 0.3641 | 0.0000 | 0.0091 |
| Completed any form of high school | 0.8220 | 0.9023 | 0.2431 | 0.0000 | 0.0148 |
| Employed with a regular high school diploma | 0.5238 | 0.7654 | 0.3754 | 0.0001 | 0.0239 |
| Employed with any high school diploma | 0.5667 | 0.7963 | 0.3737 | 0.0003 | 0.0244 |
| Never addicted or arrested | 0.3532 | 0.5444 | 0.3404 | 0.0001 | 0.0955 |

*Note*: The columns labeled *untreated mean* and *treated mean* contain the means of the participants in the control and treatment groups for the pre-program siblings (of the original Perry participants) associated with each variable (averaged at the participant-level), respectively. The column labeled *AIPW estimate* contains the augmented inverse probability weighting (AIPW) treatment effect estimates (accounting for non-response and imbalance in participants' preprogram covariates between the experimental groups). The column labeled *asymptotic p-value* contains the corresponding one-sided asymptotic $p$-value based on studentized test statistic using the analytic standard error. The column labeled *worst-case max. p* contains the worst-case maximum $p$-value based on approximate randomization tests using the studentized AIPW test statistic.

As displayed in Table 15, these effects are mostly concentrated in the sample of male pre-program siblings. The male pre-program siblings of the treated are also more likely to be never arrested or addicted compared to those related to the control group. We do not find statistically significant effects on other subgroups of pre-program siblings.

# 9 Conclusion

This paper estimates the intergenerational and intragenerational externalities of the Perry Preschool Project and finds that they are of first-order importance. Using the experimental sample of the Perry participants, we present the first experimentally based evidence on the program's spillover effects on the children and siblings of the participants. Spillover treatment effects are especially strong for the male children and siblings of the participants. We also find strong intergenerational effects on both female and male children of the male participants. Though we do not conduct a thorough mediation analysis, we note that the second-generation children in the treatment group grew up in stable two-parent homes with more parental income resources during childhood. We find intergenerational effects despite the fact that the treated families live in similar or worse neighborhoods than the untreated families. This evidence suggests that family structure is likely more important than neighborhoods in accounting for the intergenerational treatment effects on the Perry families. The program's externalities spanning multiple generations likely represent a non-negligible portion of the true rate of return to targeted high-quality preschool programs. Our findings suggest that these programs can contribute to lifting multiple generations out of poverty.

# References


Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature 56*(4), 1360–1446.

Barr, A. and C. R. Gibbs (2018). Breaking the cycle? Intergenerational effects of an anti-poverty program in early childhood. Unpublished manuscript.

Belfield, C. R. and H. M. Levin (2009). High school dropouts and the economic losses from juvenile crime in California. *California Dropout Research Project Report 16*, 1–55.

Carroll, S. J. and E. Erkut (2009). *The benefits to taxpayers from increases in students' educational attainment*, Volume 686. Rand Corporation.

Chetty, R., J. N. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2018). The opportunity atlas: Mapping the childhood roots of social mobility. *NBER Working Paper No. 25147*.

Chetty, R. and N. Hendren (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

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

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

Greenland, S. and M. A. Mansournia (2015). Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine 34*(23), 3133–3143.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic inquiry 46*(3), 289–324.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–S246.

Heckman, J. J. and G. Karapakula (2019). The Perry Preschoolers at late midlife: A study in design-specific inference. *HCEO Working Paper 2019-034*.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Huber, M. (2014). Treatment evaluation in the presence of sample selection. *Econometric Reviews 33*(8), 869–905.

Kang, J. D. Y. and J. L. Schafer (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science 22*(4), 523–539.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lunceford, J. K. and M. Davidian (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine 23*(19), 2937–2960.

Ravindran, S. (2018). Parental investments and early childhood development: Short and long run evidence from India. Unpublished manuscript.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*(427), 846–866.

Romano, J. P. and M. Wolf (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Rossin-Slater, M. and M. Wüst (2018). What is the added value of preschool? Long-term impacts and interactions with an infant health intervention. *NBER Working Paper No. 22700*.

Rumberger, R. W. and D. J. Losen (2017). The hidden costs of california's harsh school discipline: And the localized economic benefits from suspending fewer high school students. *Civil Rights Project*.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores (2005). Lifetime effects: The High/Scope Perry Preschool Study through age 40 (Monographs of the High/Scope Educational Research Foundation, 14). *Ypsilanti, MI: High Scope Educational Research Foundation*.

United States Government Accountability Office (2018). K-12 education: Discipline disparities for black students, boys, and students with disabilities. Report to congressional requesters (gao-18-258), United States Government Accountability Office.