Social Science Research 86 (2020) 102389

Contents lists available at ScienceDirect

Social Science Research

journal homepage: http://www.elsevier.com/locate/ssresearch

Heterogeneous treatment effects on Children's cognitive/non-cognitive skills: A reevaluation of an influential early childhood intervention☆

Yu Xie [a,*], Christopher Near [b], Hongwei Xu [c], Xi Song [d]

[a] Princeton University, USA
[b] Northern Arizona University, USA
[c] Queens College, City University of New York, USA
[d] University of Pennsylvania, USA

ARTICLE INFO

Keywords:
Heterogeneous treatment effect
Early child development
Perry Preschool program
Cognitive skills
Noncognitive skills

ABSTRACT

The 1962–67 High/Scope Perry Preschool Program, a well-known experimental early childhood intervention study that provided quality preschool education to disadvantaged children, has been shown to have had positive impacts on early child development and on a variety of adulthood outcomes. However, most previous analyses have only examined average treatment effects across all program participants without exploring possible effect heterogeneity by children's background characteristics. We investigated this question by first using the 1964–65 Current Population Survey data in combination with the Perry data to construct a scale of child socioeconomic status based on the estimated propensity for inclusion in the Perry program, then analyzing effect heterogeneity within the Perry sample by strata of our socioeconomic scale. We found that the treatment effects of enrollment in the Perry preschool on cognitive and non-cognitive skills were much larger and more persistent among the most disadvantaged children than among others in the Perry program. Furthermore, among the most disadvantaged children, the treatment (i.e., preschool enrollment) affects later outcomes through a reinforcement mechanism of skill development (i.e., early cognitive gain leads to a non-cognitive gain, which in turn leads to later cognitive gain) and a sequential improvement of cognitive skills over time. These findings have important implications for the evaluation of policy interventions in early child development using experimental data.

The High/Scope Perry Preschool Program was a highly effective early childhood intervention program, with robust results confirmed in many analyses of the data over the past four decades. The Perry intervention program provided preschool education to children ages 3–4 in African American families of low socioeconomic status (SES) in Ypsilanti, Michigan. A total of 123 children from 100 families were enrolled in the program in five cohorts between 1962 and 1965. Significant benefits of High/Scope preschool education have been found across a wide range of domains from early childhood to age 40 (for a review, see Schweinhart et al. 2005).

☆ Financial support for this research was provided by the National Institutes of Health, Grant R01-HD-074603-01, and by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P2CHD047879.
* Correspondence author. Department of Sociology, Princeton University, 186 Wallace Hall, Department of Sociology, Princeton University, Princeton, NJ, 08544, USA.
E-mail address: yuxie@princeton.edu (Y. Xie).

https://doi.org/10.1016/j.ssresearch.2019.102389
Received 12 March 2019; Received in revised form 4 September 2019; Accepted 12 November 2019
Available online 16 November 2019

What made the program particularly influential was its experimental design: 58 children were randomly assigned to the treatment group and 65 to the control group.

Reanalyzing the original data from the Perry Preschool Program, Heckman and his colleagues recently concluded that the main mechanism through which the Perry program benefitted the treatment subjects was the enhancement of their non-cognitive rather than cognitive skills (Heckman et al., 2013). However, we suspect that these treatment effects may be heterogeneous by subjects' likelihood of enrollment—that is, some children being more likely to be enrolled because of their low family socioeconomic status (SES)—a possibility that has not yet been examined. Indeed, almost all preschool intervention programs, ranging from regional studies, such as the Perry Preschool Program and the Abecedarian Project, to the federally funded Head Start Program, which costs about 9 billion a year (Ludwig and Phillips, 2007), have disproportionately recruited children from low SES families. If family SES is associated with unequal resource access, childrearing knowledge and expectations, cultural values, parental mental health, and parenting practices (Roubinov and Boyce, 2017), as is often the case, we should expect outcomes not only to differ between children who were enrolled and those not enrolled in these programs, but also to vary *among* children who were enrolled by their family SES. As a result, an intervention that benefits a particular subset of enrollees may not produce similar *average* treatment effects for other participants and could be incorrectly labeled ineffective.

We argue that even for experimental studies such as the Perry Preschool Program, it is useful to go beyond estimating the overall treatment effect of the treated. We already know that parents' socioeconomic status affects the propensity to enroll their children in preschool programs. We postulate that treatment effects vary by the propensity of enrollment among those who are enrolled, such that more disadvantaged children benefit more than less disadvantaged children. Understanding effect heterogeneity is important for policy evaluations of intervention programs (Manski and Garfinkel, 1992; Puma et al., 2010), especially when we wish to identify subpopulations among whom the intervention may be the most effective even if the average treatment effects among other participants are nil (Xie, 2013). In addition, knowledge about treatment effect heterogeneity helps us understand program effects for the general population or for a hypothetical population beyond the subjects being studied in an experiment. For example, preschool program effectiveness assessments often led to inconsistent findings in many past evaluation studies, but this consistency may be attributable to those studies being based on subjects varying in both socioeconomic characteristics and the likelihood of enrolling in the program. Finally, understanding treatment effect heterogeneity can also significantly improve future program designs. For these reasons, we wish to evaluate the "intervention heterogeneity hypothesis."

Our primary goal in this study is to test this intervention heterogeneity hypothesis directly. The Perry Preschool Program provides an excellent example to test our hypothesis. Inclusion of a child in the Perry study—and thus the likelihood of receiving the treatment—was based on falling below a threshold of SES; thus, we can model the likelihood of program inclusion as a function of pretreatment SES covariates. Earlier research has shown that the intervention effect for cognitive achievement is temporary, fading out in as little as four years, but none has tested whether an enduring effect may still exist among some participants with certain characteristics.

We focus on Perry Preschool subjects' likelihood of enrollment, an attribute neither directly observable nor estimable from the original data. The Perry Preschool Program only collected information on children who were enrolled, without documenting information on their peers who were excluded. To retrieve the missing information about likelihood of enrollment, we propose a method that combines experiment data with external information from population data. Specifically, we match the experimental Perry program data with the observational Current Population Survey (CPS) 1964–1965 data to create a composite dataset (Perry-CPS). The analysis strategy involves four steps: (1) using the composite samples (Perry-CPS), calculate the probability of selection into the Perry program (CPS = 0; Perry = 1) based on observed pretreatment covariates (e.g., mother's age, mother's education, mother's employment status, family structure, father's occupational skill level); (2) stratify all individuals based on their propensity of inclusion into the Perry program; (3) discard all CPS units from the composite sample (Perry-CPS), as they contain no information about treatment or outcomes; and (4) estimate the treatment effect within each propensity score stratum in the Perry sample. Note that the non-experimental CPS units are used only for modelling the likelihood of inclusion in the Perry study and not in subsequent causal analyses.

Using the composite Perry-CPS samples to estimate the propensity scores of selection into the Perry program serves two purposes. First, the propensity scores provide a summary measure of family SES to allow us to stratify the Perry sample and investigate heterogeneous treatment effects of participation in the Perry Preschool Program. Second, by comparing the Perry sample of unusually low SES African American families against an independent sample representative of the general U.S. population, the propensity score approach allows us to assess the external validity of the Perry program, i.e., the extent to which we can generalize the findings from the Perry study to the entire American population. A conventional dimension-reduction technique such as using factor analysis to construct a composite SES measure may achieve the first but not the second purpose.

We build upon a recently developed approach to identifying heterogeneous treatment effects (Xie et al., 2012) but adapt it for experimental designs in which the treatment assignment is random, but program participation may be administratively determined or self-selected. In the original Perry study, families were randomly assigned to the treatment effect (i.e., participation in the Perry program) or to a control group. Yet, participation in the study itself was not random but involved students from a sample whose members were all low in SES; one of our goals was to assess generalizability of these findings to the general population. Our research yields three new findings that modify current understanding regarding early childhood interventions. First, confirming an earlier study (Heckman et al., 2013), we show that the average treatment effect of the Perry program on cognitive skills was largely short-lived, significant at age 3–5 but inconsistent in significance at age 6 and beyond. Second, surprisingly, the Perry program had strong and persistent positive effects on children's cognitive skills measured even into early adolescence for the highest-propensity subgroup (i.e., the most disadvantaged children at baseline). Estimating overall average treatment effects without scrutinizing effect heterogeneity would overlook this important finding. Third, for this subgroup, there were two mechanisms through which the initial cognitive boost during the treatment period persisted into early adolescence. The first mechanism, proposed by Heckman and colleagues (Cunha and Heckman, 2007, 2008; Cunha et al., 2010), attends to an iterative, mutually reinforcing process between cognitive improvement and non-cognitive improvement. The second mechanism operates through a direct, sequential improvement of cognitive skills over time. Because the propensity score is essentially a composite measure of low family SES, the results indicate that children from the most socioeconomically disadvantaged families prior to the intervention benefited the most from having received the treatment. For children from other propensity score strata, there are no noticeable treatment effects on long-term cognitive or non-cognitive skill development. This finding suggests substantive and important differences in child response to treatment (i.e., participation in the Perry program) associated with the child's level of socioeconomic disadvantage. Our results may be of interest to researchers concerned with the efficacy of preschool programs, but we believe that they lead to an important methodological conclusion as well: evaluation research that does not fully account for heterogeneity in treatment effects and selection may produce ambiguous conclusions about the efficacy of the treatment.

1. Theoretical issues

1.1. *Family influences on Children's development*

A very large literature across several disciplines, including economics, psychology, sociology, education, and demography, has been devoted to understanding family influences on child development. The importance of family influences is well recognized, but scholars differ greatly as to what types of family influences are important and the mechanisms through which families exert important influences on children. Broadly speaking, scholars emphasize either monetary or nonmonetary factors as family sources of influence. In terms of mechanisms, they may emphasize children's development in either cognitive skills or non-cognitive skills, with the former referring to intellectual ability related to learning, reasoning, and problem solving, and the latter referring to soft skills, such as motivation, conscientiousness, persistence, and interpersonal interactions. We briefly summarize this literature below to provide a clear overview of why it was important to understand heterogeneous treatment effects when examining preschool outcome data. As noted above, however, our primary concern is with testing whether our proposed method for assessing heterogeneous treatment effects may produce more informative findings in evaluation research in general, not just for evaluation research focused on overall preschool outcomes.

Prominent in this literature is a well-documented association between parental income and child outcomes (Brooks-Gunn and Duncan, 1997; Duncan et al., 1998; Schneider, Hastings, and LaBriola, 2018; Yeung, Linver, and Brooks-Gunn, 2002). Parental investment theory posits that higher income allows parents to invest more money and time in educational materials (e.g., books or musical instruments), activities (e.g., visits to cultural events and school or club activities), and involvement in a child's schooling (e.g., discussing and helping with school work and attending school meetings and events), which foster child cognitive development (Mayer, 1997). Empirically, wealthier parents do spend more on such activities and materials than poorer parents (Kaushal et al., 2011) and are involved to a greater extent in school functions (Coleman, 1988; Dumais et al., 2012; Lee and Bowen, 2006; Sui-Chu and Willms, 1996). Over the recent past, although parental investment in early childhood has narrowed by some measures between rich and poor families (Bassok et al., 2016; Schneider et al., 2018), the academic achievement gap between the two has widened (Reardon, 2011).

Another explanation for the observed association between family income and child development emphasizes the causal impact of low income and material hardship (such as food insecurity, residential instability, and inadequacy of medical care) on parents' distress (Gershoff et al., 2007; Guo and Harris, 2000; Linver et al., 2002; Raver, Gershoff, and Aber, 2007; Yeung et al., 2002). High levels of stress and associated anxiety, irritability, depression, or unhealthy coping mechanisms hamper the quality of parents' interactions with children, reducing their ability to provide children with warmth, emotional support, and cognitive stimulation. This, in turn, contributes to children's depression or aggressive behaviors (Conger et al., 2010; Conger et al., 1992, 1994; Elder et al., 1992; McLeod and Shanahan, 1993).

Pushing non-monetary factors further, other scholars argue that family SES is correlated with parenting practices that affect children's outcomes. From this perspective, parenting is assumed to be associated with almost unconscious "habitus" (Bourdieu, 1977; Lewis, 1966), or more conscious and articulated values (Kohn, 1959; Kohn and Schooler, 1969) or discourses (Lareau, 2011). Whatever their form, conducive and nurturing expectations, strategies, and interactions are more commonly found among middle-class parents than among low-SES parents, and these parenting practices help children develop skills that are valued in school and later in the workplace. Several empirical studies suggest that parenting behaviors—measured either as expectations and aspirations (Bodovski and Farkas, 2008; Dumais, 2002; Gaddis, 2013; Irwin and Elley, 2011) or nurturing/stimulating practices (Cheadle 2008, 2009; Cheadle and Amato 2011; Redford et al. 2009)—are positively related to both family SES and child cognitive and behavioral skills.

1.2. *Cognitive versus non-cognitive skills in human development*

Much of the earlier research studying family influences on children focused on children's cognitive development. This is evident, for example, in the substantial literature on genetic influences of parents on children, either through twin studies or adoption studies (Bouchard and McGue, 1981; Plomin et al., 2006). Since children inherit genetic characteristics from parents, and the majority of children live with at least one parent and most with both parents, some family influences are attributable to inheritability of intelligence and personalities from parents to children. However, this genetic pathway is variable, depending on home environment

(Turkheimer et al., 2003), and is often overestimated (Devlin et al., 1997).

Besides genetic factors, research has shown that social and economic contexts also affect children's cognitive development. The importance of parents' monetary investments is widely documented, although the magnitude of this influence is still under debate. Parenting behaviors have been found to both mediate the influences of family resources and exert independent influences on children's cognitive development (Lugo-Gil and Tamis-LeMonda, 2008). Furthermore, complementing observational studies of family influences on cognitive development, several experimental intervention programs, such as the Perry High/Scope Preschool Program, the Abecedarian Project, and Infant Health and Development Studies, have also demonstrated the impact of high-quality early childcare on cognitive development immediately following intervention (Barnett, 1995).

Research interest in the role of non-cognitive skills, also known as socioemotional skills, in shaping human development has resurged in recent years (for a brief review of the history of personality psychology, see Almlund et al., 2011). Scientific evidence accumulated over the past four decades has highlighted several distinct features of non-cognitive skills that speak to their significance in potentially mediating family influences on children's socioeconomic outcomes in the long run (Hsin and Xie, 2017). Of particular importance are conscientiousness (e.g., self-control and externalizing behaviors) and neuroticism (e.g., self-esteem and locus of control). Overall, non-cognitive skills have significant effects on wage by age 30, employment, and college graduation, which is comparable to the effects of cognitive skills, and may have even stronger effects on a variety of behavioral outcomes, including daily smoking, marijuana use, marital status, teenage pregnancy (among females), and juvenile delinquency (Goldberg et al., 1998; Heckman, Stixrud, and Urzúa, 2006a; van Eijck and de Graaf, 2004).

The importance of non-cognitive skills relative to that of cognitive skills has also been found in experimental studies. In the Perry Preschool Program, the initial cognitive boost attributable to intervention among the treated participants had largely disappeared by age 10, whereas their gained advantage in non-cognitive skills—reduced externalizing behavior, improved academic motivation, and increased openness to experience—continued and contributed to better adulthood outcomes, such as higher educational attainment, higher income, and lower criminal misconduct, than those of the controlled subjects (Duncan and Magnuson, 2011).

The developments of different skills in a person are always intertwined, although researchers try to separate cognitive skills and non-cognitive skills conceptually. In reality, the development of non-cognitive skills and that of cognitive skills may well go hand-in-hand in reinforcing ways. In a series of articles, Heckman and colleagues (Cunha and Heckman, 2007, 2008; Cunha et al., 2010) formalized a model of dynamic skill formation in which "skills beget skills" through a self-reinforcing mechanism: greater skills acquired earlier in life improve the efficiency of later skill acquisition. However, they only found supporting evidence for non-cognitive skills fostering cognitive skills—at least based on empirical analyses of the Perry Preschool Program and the Children of the National Longitudinal Survey of Youth 1979—but not vice versa.

An increasing number of studies in neuroscience and genetics have also suggested possible psychosocial and epigenetic mechanisms linking environmental factors to alteration of personality traits (for a brief review, see Knudsen et al., 2006). Nevertheless, from a policy intervention perspective, a more constructive way to foster non-cognitive skills is through early-life investment by parents, schools, communities, and society at large. Existing US-based intervention studies indicate that the earlier the investment is made in the life course, the more effective and more cost-effective it will be in the long run. Overall, interventions that begin before formal schooling are more successful in boosting cognitive and non-cognitive skills and improving adulthood outcomes than those implemented in elementary school or adolescence (Kautz et al., 2014). Among early-childhood interventions that began in or before kindergarten, the improvement in cognitive skills often did not last long after the program ended, while effects on non-cognitive skills tended to persist and helped to explain variations in academic, behavioral, and socioeconomic performance in adulthood.

Despite the influential role of non-cognitive skills in human development, a remaining puzzle is how they are boosted in early interventions such as the Perry Preschool Program and sustained to exert long-lasting influences on adulthood outcomes. Cunha and Heckman (2007) suspect that the temporarily enhanced early cognitive skills that later faded out could help to foster long-term non-cognitive skills development, although no empirical evidence was reported. Knudsen et al. (2006) suggest that early experience of learning may confer value on acquired cognitive skills, which, in turn, may reinforce one's motivation and effort to learn more, that is, a reinforcing effect of cognitive skills on non-cognitive skills.

1.3. Childcare arrangement and early childhood intervention

If parenting affects children's outcomes, out-of-home childcare, as a substitute for parenting, should also matter. As documented above in our review of literature about family influences on child development, family socioeconomic status has been linked to direct and indirect associations with child outcomes, with low socioeconomic status consistently related to poor child outcomes. It then follows that early childhood intervention should particularly benefit children of low socioeconomic status. There has been a plethora of research on the effectiveness of professional early childcare using both experimental intervention studies (e.g., the Perry Preschool Program, the Abecedarian Project, and Infant Health and Development Studies) and survey-based observational studies (Belsky and Steinberg, 1978; Vandell, 2004). Most of this research suggests that childcare can have substantial benefits for children, but the mechanisms are complex and contingent on the outcomes being measured. For instance, some studies show that the long-term effects of high-quality early childcare on education, employment, health behaviors, and involvement in crime are attributable more to the improvement in non-cognitive behaviors such as self-control than that in cognitive skills. Improvements in cognitive skills, at least in experimental interventions, tended to decline and then fade completely over time (Heckman et al., 2013).

Compared to observational studies, experimental interventions are often more effective for establishing "internal validity," that is, causal conclusions within a study setting. However, many of these experimental studies on childcare arrangement have small samples that include only socioeconomically disadvantaged children from a local area, raising the question of whether their results are generalizable to a larger population. Fortunately, large-scale observational studies have yielded findings consistent with those from experimental studies and have the added advantage of showing how childcare effects vary with care provider and child characteristics. Higher quality professional childcare, associated with greater caregiver education and lower caregiver-child ratios, is associated with higher cognitive and social skills and fewer behavioral problems (Li et al., 2013; Ruzek et al., 2014; Vandell, 2004). On the other hand, in some cases (e.g., use for an extended period of time, use by children with poor social skills, or frequent changes in caregivers), professional childcare can result in greater behavioral problems and worse learning behaviors in early childhood (Barnett, 1995; Coley et al., 2013; Pilarz and Hill, 2014).

From the existing large literature, we now understand two basic facts about such early childhood intervention programs as Perry or Head Start. First, such programs are designed primarily to help children in highly disadvantaged families. An implicit corollary is that the treatment effects of the programs are heterogeneous, being most effective for children in most disadvantaged families and less effective for children in middle-class families, for the latter can provide quality in-home care or purchase professional childcare on the market. Second, given this heterogeneity in treatment effects, generalization from a local sample to any subpopulation level at large is difficult. In our analyses that follow, we do not attempt to extrapolate experimental results to a general population directly. Our goal is more modest, in revealing heterogeneous treatment effects of the Perry Preschool Program, which has direct implications if one wishes to generalize results from the Perry Preschool Program to a policy-relevant population. Our research highlights the potential usefulness in identifying special subgroups for whom an intervention would be most effective.

In addition, through reanalyzing experimental data from the Perry Preschool Program, we examine the interplay of cognitive and non-cognitive skills by uncovering a positive reinforcement process of skill formation that involves self-productivity of both cognitive and non-cognitive skills and cross-productivity between the two. We find evidence of reinforcement feedback between cognitive and non-cognitive skills development only among the most disadvantaged participants in the Perry Preschool Program. This heterogeneity in child development has been largely overlooked in the previous literature but has important academic and policy implications.

2. Data, measures, and analytic approach

2.1. Data

The primary data analyzed in the present study were drawn from the High/Scope Perry Preschool Program, with a sample of 123 African American children in Ypsilanti, Michigan, aged 3–4 at the time of enrollment in the program (1964–1967). Children were eligible for participation if they had a low Stanford-Binet IQ score (i.e., 61–89 at age 3) and low SES, measured as a function of parents' years of education, parents' occupational skill levels, and the number of rooms per person at home. Study participants were first matched in pairs based on pre-program IQ score; in each pair, one child was randomly assigned to enrollment in the Perry Preschool Program for 1–2 years, (i.e., the "treatment" effect group) and the other to a control group that received no treatment. The study then surveyed sample children annually through age 11 and by follow-up surveys at ages 14, 19, 27, and 40. Our analysis focused on the period between ages 3 and 14, when rich data on cognitive and non-cognitive skills were collected. Follow-up surveys at older ages focused on other outcomes, such as socioeconomic attainment and criminal behaviors.

We also used the nationally representative 1964–65 Current Population Survey (CPS) but restricted the cases to those comparable to the Perry data: families with an African American mother aged 18–48 with 3–12 years of education and at least one child aged 3–4 living in the periphery of an SMSA metro area similar to Ypsilanti, Michigan. Our analytical CPS sample includes 134 children for comparison, which we appended to the Perry data to produce a combined Perry-CPS sample of 257 children.

2.2. Measures

We constructed variables used in our study in two steps. First, we created the variable of likelihood of enrollment in the Perry Preschool Program using a propensity score-type method by comparing Perry school subjects with their peers in CPS samples. We used five family background variables available in both the Perry and CPS data as predictors of the propensity score: mother's education in years, mother's age in years, mother's employment status (1 = employed, 0 = not employed), family structure (1 = father present, 0 = mother only), and father's occupational skill level (1 = skilled or semi-skilled, 0 = unskilled, unemployed, or not present). We chose mother's age as a measure of children's family background because it is highly correlated with mother's age of childbearing given that all children enrolled in the program are approximately the same age. Mother's occupational skill was not included because none of the mothers had semi-skilled or skilled jobs. We chose these covariates primarily because they were the inclusion criteria for enrollment in the Perry Preschool Program.[1] Using these five covariates to predict Perry status (1 = Perry, 0 = CPS) resulted in propensity score estimates ranging from 0 to 1. We assigned an estimated propensity score to each of the subjects in the Perry program and divided them into three strata (from 1 = least likely to be enrolled or least socioeconomically disadvantaged to 3 = most likely to be enrolled and most socioeconomically disadvantaged). Note that the propensity variable refers to the likelihood of program participation, rather

[1] Parents' ages were not included among the Perry program inclusion criteria, but mother's age was a strong predictor of inclusion in our model. We did not include father's age and education because they were not strong predictors of inclusion in the Perry program in past models. Number of rooms in the house per person, another criterion for inclusion in the Perry program, was not used in our model because it was unclear which types of rooms were excluded in counts (e.g., bathrooms and garages) in each sample, and because the measure's implications for SES may not be consistent across samples because housing costs vary so greatly by region.

than the likelihood of being in the treatment group. The selection process is different from that addressed in previous work using observational data (e.g., Brand and Xie, 2010; Xie et al., 2012).

In the second step, we estimated treatment effects only in the Perry sample, for all subjects combined and by propensity score strata. We used treatment status (1 = treated, 0 = control) and propensity score stratum as our primary independent variables to predict average cognitive scores and academic motivation scores at multiple time points during and following treatment. Cognitive scores were constructed differently depending on age group. In the youngest age group (ages 3–5 years) scores from five different cognitive tests were standardized and averaged to construct an overall score during treatment: Stanford-Binet IQ scores, Leiter International Performance Scale scores, Peabody Picture Vocabulary Test III scores, and Illinois Test of Psycholinguistic Abilities (ITPA) scores. For the older age groups, we standardized and averaged scores on the California Achievement Test (CAT) for ages 6–9 (early post-treatment), and ages 10–14 (late post-treatment). We constructed academic motivation scores using annual Problem Behavior Index (PBI) items in ages 6–9 (also see Heckman et al., 2013). Teachers were asked how often the child engaged in problematic behaviors (externalizing, internalizing, social, or attention issues) on a scale from 0 ("never or rarely") through 2 ("often"). We used this index to construct two-year standardized PBI scores averaged over non-missing observations for each of two periods: early post-treatment (age 6–7) and late post-treatment (age 8–9).[2] Given the extremely low attrition throughout the Perry Preschool study,[3] we used listwise deletion and analyzed 110 Perry subjects with information on all outcome variables and key covariates.

2.3. Causal analysis

To reiterate an important point made above, "treatment effect" in this study refers to the difference between children in the treatment group of the Perry study (classified as "treated") and those who were in the control group (who were eligible but were not treated). Let Y denote an outcome, D denote treatment status (1 = treated, 0 = untreated in the Perry sample), and S denote propensity score stratum. The overall "total" treatment effect for Perry subjects is defined as:

$$TE = E[Y(D=1) - Y(D=0)] \tag{1}$$

where $Y(D=d)$ refers to counterfactual or potential outcomes, and $E(Y(D=d))$ denotes the expected outcome of Y if the treatment D is set to d. Because the Perry Preschool Program used randomization in assigning subjects to treatment status, potential outcome Y is uncorrelated with treatment status D. We call the quantity defined in Equation (1) the "total effect" because we do not control for any pre-treatment or post-treatment covariates (Manski and Garfinkel, 1992). Further, we estimated the total treatment effect by propensity as the difference between children who participated in the Perry program for 1–2 years and those who were enrolled in the Perry study as the control group within each of the estimated propensity score strata $S\ (=1,\ 2,\ 3)$ as:

$$TE_s = E(Y(D=1) - Y(D=0)|S=s) \tag{2}$$

Following Heckman et al. (2013), we repeated the analysis of cognitive and behavioral measures at different ages.[4]

As we will show later in the paper, one surprising finding from our approach is that the treatment effect of the Perry Preschool Program on cognitive scores remains significant even in early adolescence for children in the highest propensity score stratum, i.e., those from most disadvantaged families, despite its very small sample size. To interpret this effect, we use an explicit causal model, shown in Fig. 1, which involves reinforcement feedback between cognitive skills and non-cognitive skills over time (Heckman et al., 2013). We apply the model to the third propensity stratum of the Perry sample to understand why the Perry treatment effect on cognitive scores persisted into early adolescence. We aim to empirically test which of the paths posited in Fig. 1 are significant while others are not.

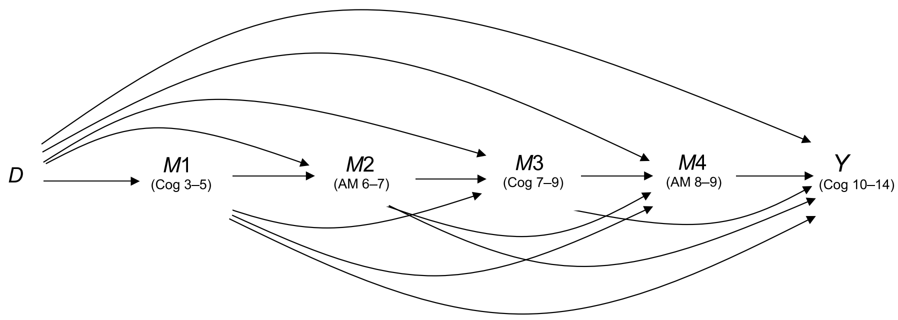
Finally, for the third stratum, we also performed a formal mediation analysis to understand the mediation mechanisms as depicted in Fig. 1. The causal mediation analysis extends the traditional Baron-Kenny (1986) mediation framework based on linear systems to nonlinear scenarios (Pearl, 2009). We define the causal mediation effect (i.e., indirect effect) of the treatment through a mediating variable M as follows:[5]

[2] Academic motivation scores were constructed using Heckman et al.'s (2013) factor loadings separately for boys and for girls, as in that study's analyses. However, the present study constructed separate academic motivation scores for each year before averaging them into two two-year increments, whereas Heckman et al. (2013) averaged PBI indicators over ages 7–9 before constructing the latent variable for that three-year range. This change was made in order to provide some temporal ordering for the academic motivation scores relative to age 7–9 cognitive scores, allowing for mediation analysis of their associations, while still providing some comparability to Heckman et al.'s (2013) construct.

[3] Only 11 out of the 123 original participants were not followed up in the age 40 interview (i.e., an attrition rate of 8.9%). Among them, four participants were missing (two in the control group and two in the treatment group), seven died (five in the control group and two in the treatment group), and no selective longitudinal attrition by treatment status was noted (Heckman et al., 2013).

[4] Note that, unlike Heckman et al. (2013), we did not conduct separate analyses for males and females. The sample size did not afford sufficient power to allow analyses by gender and by propensity score stratum simultaneously. A few previous studies have analyzed heterogeneous treatment effects of the Perry Preschool program on gender, but the results are mixed and mostly insignificant (Anderson, 2008; Schweinhart et al., 1993, 2005).

[5] Alternatively, the indirect effect of the treatment can be defined as $IE_s = E(Y(0,M(1) - Y(0,M(0))|S=s)$. These two effect estimates differ only when there is an interaction between the treatment D and the mediator M.



Notes: D = Perry School treatment; $M1$ = Mean cognitive scores at ages 3–5; $M2$ = Mean academic motivation score at ages 6–7; $M3$ = Mean cognitive scores at ages 7–9; $M4$ = Mean academic motivation score at ages 8–9; Y = Mean cognitive scores at ages 10–14.

Fig. 1. Directed Acyclic Graph (DAG) of Full Model for Perry Preschool Program Treatment Effects on Mean Cognitive Scores (10–14 Years).

$$IE_{s=3} = E(Y(D=1, M(1)) - Y(D=1, M(0)) | S = stratum\ 3), \tag{3}$$

where IE_s denotes the indirect effect of the treatment for stratum s. The indirect effect can be interpreted as the expected change in the outcome variable Y by holding the treatment status constant as $D = 1$, while changing the mediator variable from whatever values it would have attained had the treatment been changed from $D = 0$ to $D = 1$. In an experimental design, in which there is mediation through M with no confounders, Equation (3) can be estimated using the mediation formula; that is:

$$IE_s = \sum_m E(Y|D=0, M=m, S=s)[P(M|D=1, S=s) - P(M|D=0, S=s)]. \tag{4}$$

Accordingly, the direct effect of the treatment on the outcome variable that is not mediated by the intermediate variable M is defined as:

$$DE_s = E(Y(D=1, M(0)) - Y(D=0, M(0)) | S = s), \tag{5}$$

where DE_s denotes the direct effect of the treatment for stratum s. Assume that we change a child's treatment status from untreated to treated and allow all future skill development to respond to this change. The direct effect refers to changes in Y that are directly affected by this change, instead of being mediated by other associated skill changes, such as the changes in cognitive skills at age 3–5 or academic motivation at age 6–7. As discussed earlier, if such a direct effect differs for children who were in different propensity score strata, then the direct effect is heterogeneous, also known as effect modification (VanderWeele and Robins, 2007). The causal direct effect in Equation (5) can be shown to be equivalent to the controlled direct effect in experimental designs given by:[6]

$$DE_s = E(Y|D=1,\ M=m, S=s) - E(Y|D=0,\ M=m, S=s). \tag{6}$$

The estimation of total effects, either for all Perry participants as a group, or separately by propensity score strata, is not subject to the usual criticism of potential selectivity bias, as the study used randomization in assigning subjects to treatment versus control groups; however, the mediation analysis given in Equation (5) may raise this concern. The problem, of course, is that all mediators ($M1$ to $M4$) were observed outcomes and no longer randomized. Thus, we need to assume sequential ignorability in order to carry about the mediation analysis: there is no unmeasured confounder that affects both a mediator and a later mediator or the final outcome.

2.4. Sensitivity analysis

To guard against potential violation of this sequential ignorability assumption, we supplement the causal effect estimates with a sensitivity analysis that shows the extent to which the estimates are sensitive to unobserved confounders—that is, the ignorability assumption is violated. Following the bias factor approach (Arah, 2017; VanderWeele, 2015, 2016), we define the bias term for the total effect, denoted by B_S, as the product of two stratum-specific parameters:

$$B_S = \rho_S \cdot \pi_S, \tag{7}$$

[6] According to the mediation formula (Pearl, 2009), the direct effect in Equation (5) can be expressed as $DE_s = \sum_m (E(Y|D=1,\ M=m, S=s) - E(Y|D=0,\ M=m,\ S=s))P(M=m|D=0,\ S=s)$. In the absence of interaction between D and M, $E(Y|D=1,\ M=m, S=s) - E(Y|D=0,\ M=m, S=s)$ does not depend on the level of M. Therefore, the equation can be further simplified as $DE_s = E(Y|D=1,\ M=m, S=s) - E(Y|D=0,\ M=m, S=s)$.

where the two parameters are defined as:

$$\rho_S = E(Y|U=1,\ D=d,\ S=s) - E(Y|U=0,\ D=d,\ S=s\ )D=1\rho_S = E(Y|U=1,\ D=0,\ S=s) - E(Y|U=0,\ D=0,\ S=s\ ), \tag{8}$$

$$\pi_S = P(U=1|D=1, S=s) - P(U=1|D=0, S=s). \tag{9}$$

The parameter ρ refers to the mean difference in the outcome variable associated with U. The parameter π is the mean difference in the unobserved confounder U between the children who are treated and untreated in the Perry program, both conditional on the estimated propensity score stratum for entering the program. The sensitivity analysis of the mediation effect is similar to that of the total effect described above, except that both bias factors (ρ_S and π_S) need to be conditioned on the mediator, and the bias term (B_S) is equal to the negation of the product of the two parameters. For the sensitivity analyses of total effects and mediation effects, we subtract the bias term from the point estimates and the confidence intervals of the effects. The results suggest the condition of ρ_S and π_S under which the causal effects become null. If the results are very robust to unobserved confounders, we would expect the threshold values of ρ_S and π_S to be implausibly high.

3. Results

3.1. Estimating propensity score

The top panel of Table 1 shows summary statistics for the Perry and CPS samples used in the propensity score analysis. On average, families in the CPS sample were of higher SES than those in the Perry sample, with mothers' mean education of 10.7 years (compared with 9.42 years in Perry), higher percentages having an employed mother (35% versus 20%), a father present (89% versus 53%), and a father holding a skilled or semi-skilled occupation (23% versus 9%). Two-sample t-tests showed that these mean differences between the Perry and CPS samples were statistically significant.

Appendix Table A1 reports coefficient estimates from the logit model of inclusion in Perry in the combined Perry-CPS sample. All of the covariates except father's occupational skill were significant predictors of the likelihood of participation in Perry. All the coefficients (including that for father's occupational skill) had negative signs, confirming that children from higher-SES families were less likely to participate in the Perry Preschool Program. Based on the coefficient estimates, we computed the predicted propensity score of Perry program enrollment for each child. The predicted propensity scores were unsurprisingly higher on average in the Perry sample (mean = 0.62) than in the CPS sample (mean = 0.36) but still had a considerable overlap or common support given their corresponding standard deviations (0.25 in the Perry sample and 0.20 in the CPS sample; see Table 1). This suggests that the predicted propensity score is a significant, monotonic index of SES that should be controlled for our analysis of child outcomes.

The bottom panel of Table 1 presents the results of the covariate balance tests in the combined Perry-CPS sample by the three balanced propensity score strata. Constructing strata that were balanced on covariates across the two samples required using strata that were not exactly equal in relative size within or across samples. Stratum 1 consists of children who had the highest family SES and hence the lowest propensity to be included in the Perry sample. This stratum makes up about 70% of the CPS sample and 26% of the Perry sample and accounts for about 50% of the combined sample. Stratum 3 consists of the children who had the lowest family SES and hence the highest propensity to be included in Perry. This stratum makes up about 45% of the Perry sample and only 9% of the CPS sample. Nevertheless, strata 1 and 2 in the Perry sample were comparable in size (about 26% and 28%, respectively). Within each stratum, variables in the combined Perry and CPS sample were balanced according to the two-sample t-tests, which showed no statistically significant difference for all the covariates (with the exception of mother's age in stratum 3) between the two samples at the 0.05 level.

3.2. Estimating heterogeneous treatment effects of the Perry Program

Table 2 provides summary statistics for longitudinal outcome variables used in the second-step analysis of heterogeneous treatment effects of participating in the Perry program.[7] Five key longitudinal outcome variables in this study were: mean cognitive scores assessed at ages 3–5 by multiple scales (as described above) ($M1$), academic motivation at ages 6–7 as measured by PBI ($M2$), mean cognitive scores assessed at ages 7–9 by California Achievement Test (CAT) scores ($M3$), academic motivation at ages 8–9, as measured by PBI ($M4$), and mean cognitive scores assessed at ages 10–14 by CAT scores (Y). We standardized all outcome variables so that they have a mean of 0 and a standard deviation of 1 in the complete sample (i.e., before cases with missing data were removed). Within strata, between 41% and 57% of cases were treated. Means of most outcome variables were highest in stratum 1 (lowest likelihood of enrollment) between 0.14 and 0.41 and lowest in stratum 3 (highest likelihood of enrollment) between −0.33 and −0.02.

In the top panel of Table 3, we report the overall estimated total treatment effects of participating in the Perry Preschool Program on the five key outcome variables. In the entire sample, the Perry program had a positive effect of about one standard deviation on cognitive scores during the treatment period (ages 3–5) and a positive effect of about 0.39 standard deviation on cognitive scores

[7] We dropped nine observations whose propensity scores were outside the region of common support.

[8] This contrasts slightly with Heckman et al.'s (2013) finding of a positive treatment effect on academic motivation among girls that is marginally significant at the 0.05 level, due to differences in measures and analyses noted above in notes 1 and 2.

Table 1
Descriptive statistics for propensity score covariates before and after matching on socioeconomic status (SES).

	Samples				Mean Difference
	CPS		Perry		
	Mean	SD	Mean	SD	
Before matching					
Mother's education (years)	10.70	1.76	9.42	2.20	***
Mother's age (years)	31.25	5.92	29.10	6.57	**
Mother is employed (yes = 1; no = 0)	0.35	0.48	0.20	0.40	**
Father at home (yes = 1; no = 0)	0.89	0.32	0.53	0.50	***
Father in skilled occupation (yes = 1; no = 0)	0.23	0.42	0.09	0.29	**
Predicted propensity score	0.36	0.20	0.62	0.25	***
	Total $N = 134$		Total $N = 119$		
After matching					
Propensity score stratum 1 (lowest likelihood of enrollment)	$N = 89$ (72.36%)		$N = 31$ (26.05%)		
Propensity score	0.25	0.09	0.29	0.09	†
Mother's education	11.24	0.95	10.84	1.39	†
Mother's age	31.04	5.99	31.00	5.74	
Mother is employed	0.30	0.46	0.26	0.44	
Father present in home	0.99	0.11	1.00	0.00	
Father in skilled occupation	0.25	0.43	0.16	0.37	
Propensity score stratum 2 (intermediate likelihood of enrollment)	$N = 23$ (18.70%)		$N = 34$ (28.57%)		
Propensity score	0.57	0.08	0.56	0.08	
Mother's education	8.70	2.20	8.94	2.00	
Mother's age	29.57	5.91	30.65	6.98	
Mother is employed	0.30	0.47	0.21	0.41	
Father present in home	0.78	0.42	0.76	0.43	
Father in skilled occupation	0.13	0.34	0.18	0.39	
Propensity score stratum 3 (highest likelihood of enrollment)	$N = 11$ (8.94%)		$N = 54$ (45.38%)		
Propensity score	0.79	0.05	0.86	0.09	*
Mother's education	9.27	2.53	8.94	2.37	
Mother's age	31.09	4.53	27.04	6.24	*
Mother is employed	0.18	0.40	0.19	0.39	
Father present in home	0.18	0.40	0.07	0.26	
Father in skilled occupation	0.00	0.00	0.00	0.00	
	Total $N = 123$		Total $N = 119$		

Data sources: High/Scope Perry Preschool Program (1964–1967); Current Population Survey (1964–1965).
Notes: The percentage indicates the proportion of each propoensity score stratum in the CPS (or Perry sample).
†$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed tests) for significance of t-test of mean difference between CPS and Perry samples.

during the early post-treatment period (ages 7–9). However, cognitive scores among the treated subjects no longer retained significant cognitive gains during the late post-treatment period (ages 10–14). Further, perhaps surprisingly, there was no average treatment effect on academic motivation, in either age range (ages 6–7 or 8–9).[8]

When we performed the same analysis separately by the propensity score strata, reported in the lower panel of Table 3, we uncovered an important pattern in the data: for stratum 3, children from the lowest SES families, treatment effects of participating in the Perry program were stronger than in the other two groups and persisted all the way to ages 10–14, despite a reduction of the sample size by more than half. In strata 1 and 2, however, the only significant treatment effect was found for early cognitive scores ($M1$). This result confirms a gradient in treatment effects by the probability of inclusion in the Perry program.

Previous research shows that the treatment effect of participating in the Perry program on cognitive skills eventually disappeared after the Perry program ended, but the effect on non-cognitive skills remained (e.g., Cunha and Heckman, 2007; Cunha and Heckman, 2008; Cunha et al., 2010). Our findings were more complex—in agreement with the previous research in some ways but different in others—and revealed heterogeneous treatment effects across subpopulations that varied by propensity of program participation.

First, consistent with prior work, we found no significant treatment effect of participating in the Perry program on cognitive scores (as measured by the CAT scores) during the early (ages 7–9) or late post-treatment period (ages 10–14) in strata 1 and 2. However, the treatment effect on cognitive scores remained statistically significant in stratum 3 throughout the entire follow-up period, even though the effect size dropped by more than 40% from 1.13 ($M1$ in ages 3–5) to about 0.62 (Y in ages 10–14). Second, we found lingering significant treatment effects on non-cognitive skills (as measured by academic motivation) into the post-treatment period (ages 6–9), but this effect was only observed in the most disadvantaged stratum 3, not in the relatively better-off strata 1 and 2. Taken together,

[9] When effect heterogeneity depends on unobserved covariates, it is called "essential heterogeneity" by Heckman, Urzua, and Vytlacil (2006b) and "unrestrictive heterogeneity" by Deaton and Cartwright (2018).

Table 2
Descriptive statistics for treatment group and propensity score strata.

	All cases (N = 110)		Propensity Score Strata					
			Stratum 1 (lowest likelihood of enrollment; N = 30)		Stratum 2 (medium likelihood of enrollment; N = 29)		Stratum 3 (highest likelihood of enrollment; N = 51)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Treatment group (treated = 1)	0.50	0.50	0.45	0.51	0.41	0.50	0.57	0.50
Cognition at ages 3–5 (M1)								
All cases	0.00	1.00	0.14	0.98	−0.12	1.00	−0.02	1.02
Control group (treated = 0)	−0.50	0.82	−0.22	0.85	−0.58	0.87	−0.66	0.73
Treatment group (treated = 1)	0.51	0.91	0.58	0.96	0.52	0.81	0.47	0.95
Mean difference (treated - control)	1.01	***	0.80	*	1.10	**	1.13	***
Academic motivation at ages 6–7 (M2)								
All cases	0.00	1.00	0.31	0.91	−0.16	0.88	−0.09	1.09
Control group (treated = 0)	−0.14	1.01	0.42	0.66	−0.25	1.05	−0.47	1.07
Treatment group (treated = 1)	0.14	0.98	0.18	1.16	−0.03	0.57	0.19	1.04
Mean difference (treated - control)	0.27		−0.24		0.22		0.66	*
Cognition at ages 7–9 (M3)								
All cases	0.00	1.00	0.21	1.01	0.14	0.97	−0.21	0.99
Control group (treated = 0)	−0.19	0.96	0.11	0.78	−0.01	1.06	−0.57	0.90
Treatment group (treated = 1)	0.20	1.01	0.34	1.25	0.34	0.82	0.07	0.98
Mean difference (treated - control)	0.39	*	0.23		0.35		0.64	*
Academic motivation at ages 8–9 (M4)								
All cases	0.00	1.00	0.24	0.97	0.27	1.02	−0.30	0.94
Control group (treated = 0)	−0.03	1.06	0.37	0.73	0.38	1.05	−0.66	1.00
Treatment group (treated = 1)	0.03	0.95	0.09	1.21	0.12	0.99	−0.03	0.81
Mean difference (treated - control)	0.07		−0.28		−0.26		0.63	*
Cognition at ages 10–14 (Y)								
All cases	0.00	1.00	0.41	0.83	0.15	0.98	−0.33	1.01
Control group (treated = 0)	−0.09	1.07	0.47	0.73	0.14	1.14	−0.68	0.96
Treatment group (treated = 1)	0.09	0.93	0.35	0.96	0.16	0.76	−0.06	0.97
Mean difference (treated - control)	0.18		−0.11		0.01		0.62	*

Notes: This table provides the description of the longitudinal outcome variables used in the causal analysis of heterogeneous treatment effects of the Perry preschool program. All the variables are standardized so that the means are equal to zero and standard deviations are equal to 1. Roughly 50% of subjects were treated in the overall sample, but the proportions of the treated vary between 41% and 57% across propensity score strata. The means of most outcome variables are highest in stratum 1 and lowest in stratum 3, suggesting that the outcome values are positively correlated with children's propensity scores, with a higher score indicating lower SES.

Table 3
Regression estimates of total perry preschool program treatment effects on cognitive scores and academic motivation scores by propensity score stratum.

	Outcome Variables				
	Cognition at ages 3–5 (*M*1)	Academic motivation at ages 6–7 (*M*2)	Cognition at ages 7–9 (*M*3)	Academic motivation at ages 8–9 (*M*4)	Cognition at ages 10–14 (*Y*)
Total treatment effect	1.01***	0.27	0.39*	0.07	0.18
	(0.16)	(0.19)	(0.19)	(0.19)	(0.19)
Adjusted R^2	.25	.01	.03	-.01	0.00
Treatment effect by strata					
Stratum 1	0.80*	−0.24	0.23	−0.28	−0.11
(lowest likelihood of enrollment)	(0.31)	(0.35)	(0.35)	(0.34)	(0.35)
Stratum 2	1.10**	0.22	0.35	−0.26	0.01
(intermediate likelihood of enrollment)	(0.33)	(0.37)	(0.37)	(0.36)	(0.36)
Stratum 3	1.13***	0.66*	0.64*	0.63*	0.62*
(highest likelihood of enrollment)	(0.25)	(0.28)	(0.28)	(0.27)	(0.27)
Adjusted R^2	.24	.05	.05	.09	.11

Notes: $N = 110$. The total effect of the Perry Preschool Program on any standardized outcome variable M or Y is defined as the mean difference in Y between the groups of the treated ($D = 1$) and the untreated cases ($D = 0$). Standard errors are shown in parentheses. Treatment effect denotes the average difference in standardized outcome between treated cases and untreated cases in a given propensity score stratum.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed tests).

these findings suggested heterogeneity in the long-term effect of the Perry program, with the most disadvantaged children benefiting the most. By focusing exclusively on the overall treatment effects, previous studies have overlooked this heterogeneity and over-simplified the long-term implications of the early childhood intervention. In addition, the significant effects on both cognitive and non-cognitive scores during the post-treatment period in stratum 3 suggests the plausibility of a model of skill formation that illustrates how "skills beget skills." We test this dynamic process more formally in the next section.

3.3. Mediation analyses

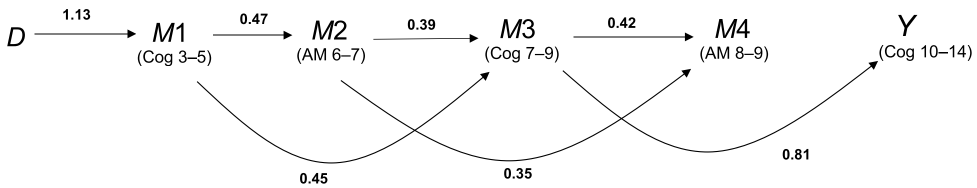
How did the Perry initial treatment on the 3-4-year-olds produce long-term beneficial effects for children in the lowest SES families (stratum 3)? In Fig. 1, we provide a directed acyclic graph (DAG) to hypothesize this process, drawing from the theoretical insight that cognitive skills and non-cognitive skills may reinforce each other over the life course. In this model, we use the cognitive scores at ages 10–14 as the final outcome variable (Y), with variables $M1$–$M4$ as mediating variables. The model is conservative, as it allows for all possible pathways linking any given preceding cognitive or non-cognitive measure to a later measure. We estimated regression models that included all of these possible pathways (results shown in Appendix Table A2) and a series of more parsimonious models (not shown). We also examined models that included interaction terms but most of these coefficients were not significant, as expected, due to the small sample size.

Many paths in the model were not statistically different from zero. This is not surprising given the limited statistical power resulting from a small sample size ($N = 51$). Thus, to obtain more reliable estimates for the paths between mediating variables and to simplify the mediation analysis, we trimmed paths that were not significantly different from zero and obtained a simplified model that balanced model accuracy with parsimony. Estimates for this simplified model are summarized graphically in Fig. 2. We chose this simplified model mainly for the sake of being parsimonious in light of weak statistical power, but not because we assume the absence of other potential paths in theory.

The full results are presented in Table 4, with coefficients in a series of models that regressed each post-treatment assessment on its most recent prior measure of cognitive and non-cognitive skills. Cognitive scores at ages 3–5 (during the treatment period and denoted by $M1$) were positively associated with academic motivation at ages 6–7 (denoted by $M2$), suggesting that the boost to cognitive skills through the Perry program helped improve children's non-cognitive skills even after the intervention ended. Cognitive scores at ages 3–5 were also positively associated with cognitive scores at ages 7–9 (denoted by $M3$), as was academic motivation at ages 6–7. These results suggest that the treatment-induced cognitive improvement could have a lasting impact on cognitive achievement in the early post-treatment period through both a direct pathway and an indirect pathway—by inspiring academic motivation at ages 6–7. Similar patterns held for academic motivation at ages 8–9 (denoted by $M4$) and CAT score at ages 10–14 (denoted by Y)—they were positively predicted by the most recent prior measures of cognitive and non-cognitive skills. The only exception was that academic motivation at ages 8–9 was not significantly associated with cognitive scores at ages 10–14. This finding may imply that growth of some non-cognitive skills does not continue to stimulate growth of cognitive skills beyond a certain developmental stage. As an alternative, however, this finding may simply reflect the fact that we do not have sufficient statistical power due to the small sample size.

The simplified pathways in Fig. 2 for stratum 3 revealed two distinct mechanisms through which the Perry program could have an enduring impact on cognitive achievement long after the intervention ended. The first mechanism corresponds to the dynamic model that we discussed earlier, which involves a recursive and mutually reinforcing process between improvement in cognitive and non-cognitive skills. Specifically, the treated children received a cognitive boost during the treatment period (represented by the straight arrow pointing from D to $M1$); the early cognitive improvement stimulated greater academic motivation at ages 6–7 (the straight arrow pointing from $M1$ to $M2$) which, in turn, led to significantly higher later cognitive scores at ages 7–9 (the straight arrow pointing from $M2$ to $M3$); the sustained cognitive advantage positively reinforced academic motivation through ages 8–9 (the straight arrow pointing from $M3$ to $M4$).

The second mechanism corresponds to sequential improvement of cognitive skills, which was set in motion by the initial boost in cognitive scores caused by the Perry treatment. Again, the treated children in the Perry program gained a cognitive boost during the treatment period; the initial advantage was then converted to a subsequent advantage at ages 7–9 (the curved arrow pointing from $M1$



Notes: D = Perry School treatment; $M1$ = Mean cognitive scores at ages 3–5; $M2$ = Mean academic motivation score at ages 6–7; $M3$ = Mean cognitive scores at ages 7–9; $M4$ = Mean academic motivation score at ages 8–9; Y = Mean cognitive scores at ages 10–14. Only coefficients statistically significant at the $p < .05$ level are shown.

Fig. 2. Directed Acyclic Graph (DAG) of Results from Trimmed Model (Model B) Regression Analyses for Perry Preschool Program Treatment Effect on Mean Cognitive Scores (10–14 Years) for Propensity Score Stratum 3.

Table 4
Regression estimates of direct effects on academic motivation scores and cognitive scores for propensity score stratum 3 (lowest SES).

	Outcome Variables			
	Academic motivation at ages 6–7 (*M*2)	Cognition at ages 7–9 (*M*3)	Academic motivation at ages 8–9 (*M*4)	Cognition at ages 10–14 (*Y*)
Cognition at ages 3–5 (*M*1)	0.47*** (0.12)	0.45*** (0.10)		
Cognition at ages 7–9 (*M*3)			0.42*** (0.12)	0.81*** (0.10)
Academic motivation at ages 6–7 (*M*2)		0.39*** (0.10)	0.35** (0.10)	
Academic motivation at ages 8–9 (*M*4)				0.14 (0.11)
R^2	0.24	0.58	0.64	0.76
Adjusted R^2	0.21	0.55	0.61	0.74

Notes: $N = 110$. According to Equation (5), the causal estimates of the direct effect can be estimated by the controlled direct effect so that $DE_s = E(Y|D = 1,\ M = m, S = s) - E(Y|D = 0,\ M = m, S = s)$. The row variables (*M*1, *M*2, *M*3, *M*4) refer to a sequence of treatments. The column variables (*M*2, *M*3, *M*4, and *Y*) refer to a sequence of outcome variables. Standard errors are shown in parentheses.
*$p < 0.05$. **$p < 0.01$, ***$p < 0.001$ (two-tailed tests).

to *M*3), which in turn affected the subsequent cognitive development in ages 10–14 (the curved arrow pointing from *M*3 to *Y*).

In Table 5, we present the main results of the mediation analysis for stratum 3. It shows that cognitive scores in ages 3–5 (*M*1) served as a significant mediator for academic motivation in ages 6–7 (*M*2) and cognitive scores in ages 7–9 (*M*3). Academic motivation in ages 6–7 also mediated the treatment effects of participating in the Perry program on cognitive scores in ages 7–9 and academic motivation in ages 8–9 (*M*4). Finally, cognitive scores in ages 7–9 mediated the treatment effects on both academic motivation in ages 8–9 and cognitive scores in ages 10–14 (*Y*). We did not find a mediation effect of academic motivation in ages 8–9 on cognitive scores in ages 10–14.

3.4. Mediation sensitivity analysis

How are the mediation results sensitive to the sequential ignorability assumption? Appendix Table B1 shows sensitivity analysis results for a sequence of mediating effects for stratum 3. We assumed a binary confounding variable, $U = 0$ or 1, for each subject, and varied the values of ρ from -2 to 2, suggesting that each one-unit difference in U would lead to ρ unit changes in the mean score of children's outcome variable. Because all of the outcome variables were standardized, the unit change in ρ could also be interpreted as change in standard deviation. We fixed the value of π at -10 percent and -20 percent because most pretreatment characteristics of children in the treatment and control groups differed within 10 percent and very few by more than 20 percent. The results in Table B1 suggested that the mediating effect of *M*1 on *M*3 was very robust to the presence of unobserved confounding, even if the effect of the hypothetical confounder *U* on *M*3 was as large as two standard deviations. For the rest of the mediators, omitting the effect of a potential confounder could lead to overestimated effects of prior cognitive ability and prior academic motivation. Yet, the mediation effects remained significant if π was fixed at -10 percent and ρ was higher than -0.5 (i.e., the expected outcome difference with and without the confounder is smaller than 0.5 standard deviation).

Table 5
Heterogeneous mediation effects of perry preschool program treatment (D) on academic motivation measures and cognitive scores for propensity score stratum 3.

	Outcome Variables							
	Academic motivation at ages 6–7 (*M*2)		Cognition at ages 7–9 (*M*3)		Academic motivation at ages 8–9 (*M*4)		Cognition at ages10–14 (*Y*)	
Mediating Variables	Mediation effect	Proportion mediated	Mediation effect	Proportion mediated	Mediation effect	Proportion mediated	Mediation effect	Proportion mediated
Cognition at ages 3–5 (*M*1)	0.50* (0.21)	74%	0.77** (0.22)	116%				
Cognition at ages 7–9 (*M*3)					0.41* (0.19)	62%	0.58* (0.25)	90%
Academic motivation at ages 6–7 (*M*2)			0.36* (0.18)	55%	0.37* (0.18)	57%		

Notes: $N = 110$. Standard errors are shown in parentheses. We have estimated models with a full set composed of the treatment variable and all mediating variables, but the results here show only the final models that include statistically significant variables. Only significant coefficients are included in our final models. If there is more than one mediating variable (for example, M1 and M2 both mediate the effect of D on M3), we estimate two separate models to estimate the mediation effect of a single mediator while controlling for the effect of the other as a fixed covariate.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed tests).

4. Discussion and conclusion

Decades of social research have provided clear evidence that intergenerational transmission of social advantages and disadvantages occurs early in a child's life (Blau and Duncan, 1967; Bourdieu, 1977; Duncan and Brooks-Gunn, 1997; Hauser et al., 1983; Mayer, 1997; McLanahan and Sandefur 1994; Sewell et al., 1969). An increasing body of scholarship has pointed to the importance of early childhood in support of preschool intervention programs targeting children in socially disadvantaged families. Head Start, funded by the federal government, is one of the best known and the most expensive of these programs.

Yet it is difficult to evaluate the effectiveness of these programs, in part because individuals who are enrolled in a program may be different from those who are not in terms of observed, as well as unobserved, characteristics. To overcome methodological shortcomings of observational data, experimental intervention programs such as the Perry Program and the Abecedarian Project have been conducted to evaluate the positive benefits of high-quality early child intervention for disadvantaged children—namely, the subjects in these studies. It is widely believed that experimental designs can be an effective solution to the selection bias problem: randomization guarantees that the only difference between subjects in the treatment versus control group is attributed to treatment. We argue that this is not necessarily the case and that the experimental method may have its methodological limitations as well. The problem, of course, lies in the difficulty of generalizing estimates obtained from a typically unrepresentative sample to any meaningful population (see Deaton and Cartwright, 2018 for a discussion). In particular, if treatment effects are heterogeneous, generalizing results from an experimental setting to a general population is problematic and indeed impossible without making strong and untestable assumptions (Deaton and Cartwright, 2018; Manski and Garfinkel, 1992; Stuart et al., 2011; Subramanian et al., 2018).

If treatment effect heterogeneity depends only on observed covariates, a situation termed "restrictive heterogeneity" by Deaton and Cartwright (2018),[9] one can apply appropriate weights in order to generalize results to broader samples or populations. However, estimating effect variability with high-dimensional covariates suffers from the curse of dimensionality. As Xie et al. (2012) pointed out, given ignorability, which is satisfied by randomization, the question remains as to whether results from a particular study are generalizable to the whole or any other population. In this situation, the evidence of consequential treatment effect of heterogeneity depends directly on the propensity of treatment. Stuart et al. (2011) also emphasized the key role of the propensity score in generalizing results from experimental data to population-level estimands. We implemented these previous methods by merging experimental data with external, population-representative data in order to assess heterogeneous treatment effects by the propensity score within an experimental study.

In this study, we sought to examine heterogeneous effects by using the Perry Study to identify three strata of socioeconomic background. As noted above, the Perry Study was an intervention of low-income families, so the range of family socioeconomic background in the Perry Study was compressed by construct (e.g., Table 1 shows that families in the highest category of SES in the Perry Study are still significantly less socioeconomically advantaged than the CPS sample). Thus, we seek to identify heterogeneous effects within a sample of families who are already very disadvantaged. We speculate that our findings have implications for the population from which this sample is drawn (i.e., all preschool children) but we need to be cautious in generalizing from our results to samples with greater variance in SES. Yet it is important to note that many studies of preschool program outcomes focus on low SES samples that are at least somewhat more similar to the Perry sample; we believe that our findings may generalize to these low SES samples, although differences may still arise (e.g., due to area of the US from which the sample was drawn).

Generalization of results from experimental data is a methodological subject beyond the scope of this paper, in fact. From our study, one substantive contribution is to show empirically that the presence of treatment heterogeneity may obscure the effect of an influential randomized intervention—the Perry Preschool Program. Many scholars and policymakers believe that we should provide policy interventions targeting "disadvantaged" children. However, how do we define "disadvantaged?" What samples are best targeted with early childhood policy interventions, in order to gain the greatest benefits?

We have shown that the meaning of disadvantage can be further explored through examining heterogeneity within experimental data—data from the Perry Preschool Program in our study. By comparing subjects who participated in the Perry study to others who did not but were observed in the Current Population Survey, we were able to rank-order and then classify the Perry subjects by their propensity to participate in the study. Since the propensity score was a composite of pretreatment covariates, stratification by the propensity score could not introduce bias to the analysis. That is, we still retained the validity of the results due to the original experimental design, albeit for subsamples defined by the propensity score.

We found that the Perry Preschool Program produced the largest and most long-lasting causal effects in the stratum most likely to participate in the program (i.e., children in the most disadvantaged families). For Perry subjects living in the most disadvantaged families, we found evidence in support of a dynamic model proposed by Heckman (2006) and Heckman et al. (2013), which posits a feedback cycle between cognitive development and non-cognitive development. The Perry Preschool Program produced an immediate benefit to the cognitive scores of the preschool-aged children who were treated. The higher cognitive scores of the treated children boosted their confidence and interest in academic studies, which in turn improved their subsequent cognitive scores. In addition to this mediating mechanism of non-cognitive skills, we also found a direct path from an early cognitive gain to a later cognitive gain. Again, this finding would not have surfaced had we not conducted an analysis limited to the children from the most disadvantaged families.

While we found most treatment effects to be statistically significant for the last stratum but not for other strata, we had insufficient statistical power with which to test whether treatment effects were statistically different across strata, except for the two motivation outcomes (Appendix Table A3, Panel B). Also, recall that the propensity score was a composite of component covariates. Treatment effect by the propensity score thus implies interaction effects between treatment and component covariates. In Panel A of Appendix Table A3, we show statistical significance of interaction effects between treatment and covariates. We observed that among pretreatment covariates the treatment effects of the Perry program differed significantly only by mother's education: the term for the interaction between treatment and mother's education had an inverse association with academic motivation scores in both age ranges. These results suggest that mother's education level was a major driving force behind the heterogeneous treatment effects we uncovered in the Perry Preschool Program, with children of less educated mothers benefiting more than children of more educated mothers. Research on parenting practices has long shown that more educated mothers had greater educational expectations and aspirations for their children than did less educated mothers, resulting in greater academic motivation among their children, as compared to other children. The Perry Preschool Program data show that an external intervention helped bridge this gap in motivation by encouraging disadvantaged children, which included (among other characteristics) children of less-educated mothers. This suggests that other early childhood education programs may similarly be most effective at instilling positive approaches to learning and building non-cognitive skills among the most disadvantaged children.

Human beings are inherently heterogeneous. Not only do they differ from one another in numerous characteristics that may be easily observed by others, including social scientists, and many more that are unobserved; they also differ in responses to the same treatments or stimuli experienced in reality or in real-life experiments. Public policies are often designed with this population heterogeneity in mind, targeting an intervention program to individuals who would benefit most from it. However, we should pay attention not only to between-group differences— say individuals enrolled or not enrolled in a program—but also within-group variability—how subjects who are all enrolled in a program under some recruitment criteria may still be different from each other.

By combining external data—from the Current Population Survey—with experimental data from the Perry Preschool Program, we found substantial heterogeneity: the treatment effects of Perry are more pronounced and more persistent among the subjects in the most disadvantaged stratum than among the other Perry subjects, who were also disadvantaged. Furthermore, among these most disadvantaged children, the treatment affects later outcomes through a reinforcement mechanism of skill development (i.e., early cognitive gain leads to non-cognitive gain, which in turn leads to later cognitive gain) and a sequential improvement of cognitive skills over time. However, we also note that our study is limited by data from a single experimental program that had a small sample size and was carried out more than five decades ago. We welcome future research to replicate and challenge the results presented in the study.

Appendix A. Detailed Statistical Results

Table A1
Logit Model Estimates for Propensity of Being Selected for Perry Sample (0 = CPS, 1 = Perry).

Covariates	b	Standard error
Mother's education (years)	−0.21***	0.05
Mother's age (years)	−0.04*	0.02
Mother is employed (yes = 1; no = 0)	−0.54**	0.21
Father present in home (yes = 1; no = 0)	−1.24***	0.22
Father is employed in skilled occupation (yes = 1; no = 0)	−0.09	0.26
Constant	4.31***	0.75

Notes:*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tailed tests. Likelihood-ratio chi-squared (5 degrees of freedom) = 81.79; Pseudo $R^2 = 0.23$. $N = 253$.

Table A2
OLS Regression Models of Treatment and Mediating Variables on Later Variables, by Propensity Score Stratum

	Outcome Variables			
	Academic motivation at ages 6–7 (*M*2)	Cognition at ages 7–9 (M3)	Academic motivation at ages 8–9 (M4)	Cognition at ages 10–14 (Y)
Stratum 1				
Treatment (*D*)	−0.71*	0.21	−0.48†	0.19
	(0.35)	(0.30)	(0.26)	(0.24)
Cognition at ages 3–5 (*M*1)	0.59**	0.23	0.21	−0.17
	(0.18)	(0.17)	(0.16)	(0.14)
Academic motivation at ages 6–7 (*M*2)		0.70***	0.32†	0.35*
		(0.17)	(0.19)	(0.17)
Cognition at ages 7–9 (*M*3)			0.49**	0.16
			(0.17)	(0.17)
Academic motivation at ages 8–9 (*M*4)				0.36†
				(0.21)
Constant	0.54*	−0.13	0.23	0.11
	(0.22)	(0.19)	(0.17)	(0.15)
Stratum 2				
Treatment (*D*)	−0.39	0.06	−0.48†	0.06
	(0.40)	(0.31)	(0.28)	(0.24)
Cognition at ages 3–5 (*M*1)	0.55**	0.13	−0.06	−0.11
	(0.20)	(0.18)	(0.16)	(0.13)
Academic motivation at ages 6–7 (*M*2)		0.70***	−0.19	0.50**
		(0.17)	(0.19)	(0.16)
Cognition at ages 7–9 (*M*3)				

(continued on next page)

Table A2 (*continued*)

	Outcome Variables			
	Academic motivation at ages 6–7 (*M*2)	Cognition at ages 7–9 (M3)	Academic motivation at ages 8–9 (M4)	Cognition at ages 10–14 (Y)
			0.94***	0.21
			(0.17)	(0.19)
Academic motivation at ages 8–9 (*M*4)				0.42**
				(0.14)
Constant	0.06	0.24	0.30†	0.05
	(0.24)	(0.19)	(0.17)	(0.15)
Stratum 3				
Treatment (*D*)	0.17	−0.18	0.37†	−0.07
	(0.30)	(0.23)	(0.21)	(0.17)
Cognition at ages 3–5 (*M*1)	0.43**	0.50***	−0.28*	0.10
	(0.15)	(0.12)	(0.12)	(0.11)
Academic motivation at ages 6–7 (*M*2)		0.39***	0.33**	0.01
		(0.10)	(0.10)	(0.09)
Cognition at ages 7–9 (*M*3)			0.55***	0.73***
			(0.13)	(0.13)
Academic motivation at ages 8–9 (*M*4)				0.17
				(0.12)
Constant	−0.19	−0.06	−0.37*	−0.08
	(0.21)	(0.16)	(0.14)	(0.13)
Adjusted R^2	.23	.54	.64	.76

Notes: $N = 110.$, †$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; two-tailed tests.

Table A3
Significance of Heterogeneous Treatment Effects by Covariates and Propensity Score Stratum

	Outcome Variables				
	Cognition at ages 3–5 (*M*1)	Academic motivation at ages 6–7 (*M*2)	Cognition at ages 7–9 (*M*3)	Academic motivation at ages 8–9 (*M*4)	Cognition at ages 10–14 (*Y*)
Panel A: Covariates					
Mother's education		*		*	
Mother's age					
Mother works					
Father present in home		†			
Father in skilled occupation					
Panel B: Propensity Stratum					
Stratum 1 versus 2					
Stratum 1 versus 3		*		*	†
Stratum 2 versus 3				*	

Notes: N = 110., †p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001; two-tailed tests.

Appendix B. Mediation Analyses

Table B1
Mediation Effects of Perry Preschool Program Treatment (D) on Academic Motivation Score and Cognitive Scores for Propensity Score Stratum 3

Mediating Variables	Sensitivity parameters		Outcome Variables							
			Academic motivation at ages 6–7 (*M*2)		Cognition at ages 7–9 (*M*3)		Academic motivation at ages 8–9 (*M*4)		Cognition at ages 10–14 (*Y*)	
	ρ_s	π_s	Mediation effect (95% CI)		Mediation effect(95% CI)		Mediation effect(95% CI)		Mediation effect (95% CI)	
Cognition at ages 3–5 (*M*1)	−2.00	−10%	0.30	(-0.11, 0.71)	0.57	(0.14, 1.00)	–	–	–	–
	−1.50	−10%	0.35	(-0.06, 0.76)	0.62	(0.19, 1.05)	–	–	–	–
	−1.00	−10%	0.40	(-0.01, 0.81)	0.67	(0.24, 1.10)	–	–	–	–
	−0.50	−10%	0.45	(0.04, 0.86)	0.72	(0.29, 1.15)	–	–	–	–
	0.50	−10%	0.55	(0.14, 0.96)	0.82	(0.39, 1.25)	–	–	–	–
	1.00	−10%	0.60	(0.19, 1.01)	0.87	(0.44, 1.30)	–	–	–	–
	1.50	−10%	0.65	(0.24, 1.06)	0.92	(0.49, 1.35)	–	–	–	–

(*continued on next page*)

Table B1 (*continued*)

Mediating Variables	Sensitivity parameters		Outcome Variables							
			Academic motivation at ages 6–7 (*M*2)		Cognition at ages 7–9 (*M*3)		Academic motivation at ages 8–9 (*M*4)		Cognition at ages 10–14 (*Y*)	
	ρ_s	π_s	Mediation effect (95% CI)		Mediation effect(95% CI)		Mediation effect(95% CI)		Mediation effect (95% CI)	
	2.00	−10%	0.70	(0.29, 1.11)	0.97	(0.54, 1.40)	–	–	–	–
Cognition at ages 7–9 (*M*3)	−2.00	−10%	–	–	–	–	0.21	(-0.17, 0.58)	0.38	(-0.11, 0.87)
	−1.50	−10%	–	–	–	–	0.26	(-0.12, 0.63)	0.43	(-0.06, 0.92)
	−1.00	−10%	–	–	–	–	0.31	(-0.07, 0.68)	0.48	(-0.01, 0.97)
	−0.50	−10%	–	–	–	–	0.36	(-0.02, 0.73)	0.53	(0.04, 1.02)
	0.50	−10%	–	–	–	–	0.46	(0.08, 0.83)	0.63	(0.14, 1.12)
	1.00	−10%	–	–	–	–	0.51	(0.13, 0.88)	0.68	(0.19, 1.17)
	1.50	−10%	–	–	–	–	0.56	(0.18, 0.93)	0.73	(0.24, 1.22)
	2.00	−10%	–	–	–	–	0.61	(0.23, 0.98)	0.78	(0.29, 1.27)
Academic motivation at ages 6–7 (*M*2)	−2.00	−10%	–	–	0.16	(-0.19, 0.51)	0.17	(-0.18, 0.52)	–	–
	−1.50	−10%	–	–	0.21	(-0.14, 0.56)	0.22	(-0.13, 0.57)	–	–
	−1.00	−10%	–	–	0.26	(-0.09, 0.61)	0.27	(-0.08, 0.62)	–	–
	−0.50	−10%	–	–	0.31	(-0.04, 0.66)	0.32	(-0.03, 0.67)	–	–
	0.50	−10%	–	–	0.41	(0.06, 0.76)	0.42	(0.07, 0.77)	–	–
	1.00	−10%	–	–	0.46	(0.11, 0.81)	0.47	(0.12, 0.82)	–	–
	1.50	−10%	–	–	0.51	(0.16, 0.86)	0.52	(0.17, 0.87)	–	–
	2.00	−10%	–	–	0.56	(0.21, 0.91)	0.57	(0.22, 0.92)	–	–
Cognition at ages 3–5 (*M*1)	−2.00	−20%	0.10	(-0.31, 0.51)	0.37	(-0.06, 0.80)	–	–	–	–
	−1.50	−20%	0.20	(-0.21, 0.61)	0.47	(0.04, 0.90)	–	–	–	–
	−1.00	−20%	0.30	(-0.11, 0.71)	0.57	(0.14, 1.00)	–	–	–	–
	−0.50	−20%	0.40	(-0.01, 0.81)	0.67	(0.24, 1.10)	–	–	–	–
	0.50	−20%	0.60	(0.19, 1.01)	0.87	(0.44, 1.30)	–	–	–	–
	1.00	−20%	0.70	(0.29, 1.11)	0.97	(0.54, 1.40)	–	–	–	–
	1.50	−20%	0.80	(0.39, 1.21)	1.07	(0.64, 1.50)	–	–	–	–
	2.00	−20%	0.90	(0.49, 1.31)	1.17	(0.74, 1.60)	–	–	–	–
Cognition at ages 7–9 (*M*3)	−2.00	−20%	–	–	–	–	0.01	(-0.37, 0.38)	0.18	(-0.31, 0.67)
	−1.50	−20%	–	–	–	–	0.11	(-0.27, 0.48)	0.28	(-0.21, 0.77)
	−1.00	−20%	–	–	–	–	0.21	(-0.17, 0.58)	0.38	(-0.11, 0.87)
	−0.50	−20%	–	–	–	–	0.31	(-0.07, 0.68)	0.48	(-0.01, 0.97)
	0.50	−20%	–	–	–	–	0.51	(0.13, 0.88)	0.68	(0.19, 1.17)
	1.00	−20%	–	–	–	–	0.61	(0.23, 0.98)	0.78	(0.29, 1.27)
	1.50	−20%	–	–	–	–	0.71	(0.33, 1.08)	0.88	(0.39, 1.37)
	2.00	−20%	–	–	–	–	0.81	(0.43, 1.18)	0.98	(0.49, 1.47)
Academic motivation at ages 6–7 (*M*2)	−2.00	−20%	–	–	−0.04	(-0.39, 0.31)	−0.03	(-0.38, 0.32)	–	–
	−1.50	−20%	–	–	0.06	(-0.29, 0.41)	0.07	(-0.28, 0.42)	–	–
	−1.00	−20%	–	–	0.16	(-0.19, 0.51)	0.17	(-0.18, 0.52)	–	–
	−0.50	−20%	–	–	0.26	(-0.09, 0.61)	0.27	(-0.08, 0.62)	–	–
	0.50	−20%	–	–	0.46	(0.11, 0.81)	0.47	(0.12, 0.82)	–	–
	1.00	−20%	–	–	0.56	(0.21, 0.91)	0.57	(0.22, 0.92)	–	–
	1.50	−20%	–	–	0.66	(0.31, 1.01)	0.67	(0.32, 1.02)	–	–
	2.00	−20%	–	–	0.76	(0.41, 1.11)	0.77	(0.42, 1.12)	–	–

Notes: ρ_s refers to the mean difference in children's outcomes associated with a unit difference in the binary unobserved confounder conditional on treatment status, propensity scores, and mediator values. π_s refers to the prevalence difference of the binary unobserved confounder between children of treatment and control groups conditional on treatment status, propensity scores, and mediator values. The bias factor is equal to the negation of the product of the two parameters, and we subtract this bias factor from the mediation effect and the confidence interval. We omit sensitivity analyses for all mediators that are not statistically significant in Table 5. The mediation effects (i.e., indirect effects) are defined in Equation (5).

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssresearch.2019.102389.

References

Almlund, Mathilde, Lee Duckworth, Angela, Heckman, James J., Kautz, Tim, 2011. Personality Psychology and Economics. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, vol. 4. Elsevier, North-Holland, Amsterdam. Pp. 1–181.

Anderson, Michael L., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, Perry preschool, and early training projects. J. Am. Stat. Assoc. 103 (484), 1481–1495.

Arah, Onyebuchi, 2017. Bias analysis for uncontrolled confounding in the health science. Annu. Rev. Public Health 38, 12.1–12.16.

Barnett, Steven W., 1995. Long-Term effects of early childhood programs on cognitive and school outcomes. Future Child. 5, 25–30.

Baron, Reuben M., Kenny, David A., 1986. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Personal. Soc. Psychol. 51 (6), 1173–1182.

Bassok, Daphna, Finch, Jenna, Lee, RaeHyuck, Reardon, Sean F., Jane, Waldfogel, 2016. Socioeconomic gaps in early childhood experiences: 1998 to 2010. AERA Open 2 (3), 1–22.

Belsky, Jay, Steinberg, Laurence D., 1978. The effects of day care: a critical review. Child Dev. 49 (4), 929–949.

Blau, Peter, Duncan, Otis Dudley, 1967. The American Occupational Structure. John Wiley & Sons, New York, NY.

Bodovski, Katerina, Farkas, George, 2008. 'concerted cultivation' and unequal achievement in elementary school. Soc. Sci. Res. 37, 903–919.

Bouchard Jr., Thomas J., McGue, Matthew, 1981. Familial studies of intelligence: a review. Science 212, 1055–1059.

Bourdieu, Pierre, 1977. Cultural reproduction and social reproduction. Pp. 487–511. In: Karabel, J., Halsey, A.H. (Eds.), Power and Ideology in Education. Oxford University Press, New York, NY.

Brand, Jennie E., Xie, Yu, 2010. Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. Am. Sociol. Rev. 75 (2), 273–302.

Brooks-Gunn, Jeanne, Duncan, Greg J., 1997. The effects of poverty on children. Future Child. 7, 55–71.

Cheadle, Jacob E., 2008. Educational investment, family context, and children's math and reading growth from Kindergarten through the third grade. Sociol. Educ. 81, 1–31.

Cheadle, Jacob E., 2009. Parent educational investment and children's general knowledge development. Soc. Sci. Res. 38, 477–491.

Cheadle, Jacob E., Amato, Paul R., 2011. A quantitative assessment of lareau's qualitative conclusions about class, race, and parenting. J. Family Issues 32, 679–706.

Coleman, James S., 1988. Social capital in the creation of human capital. Am. J. Sociol. 94, S95–S120.

Coley, Rebekah Levine, Votruba-Drzal, Elizabeth, Miller, Portia L., Koury, Amanda, 2013. Timing, extent, and type of child care and children's behavioral functioning in kindergarten. Dev. Psychol. 49, 1859–1873.

Conger, Rand D., Conger, Katherine J., Elder Jr., Glen H., Lorenz, Frederick O., Simons, Ronald L., Whitbeck, Les B., 1992. A family process model of economic hardship and adjustment of early adolescent boys. Child Dev. 63, 526–541.

Conger, Rand D., Ge, Xiaojia, Elder Jr., Glen H., Lorenz, Frederick O., Simons, Ronald L., 1994. Economic stress, coercive family process, and developmental problems of adolescents. Child Dev. 65, 541–561.

Conger, Rand D., Conger, Katherine J., Martin, Monica J., 2010. Socioeconomic status, family processes, and Individual Development. J. Marriage Family 72, 685–704.

Cunha, Flavio, Heckman, James J., 2007. The technology of skill formation. Am. Econ. Rev. 97, 31–47.

Cunha, Flavio, Heckman, James, 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. J. Hum. Resour. 43, 738–782.

Cunha, Flavio, Heckman, James J., Schennach, Susanne M., 2010. Estimating the technology of cognitive and noncognitive skill formation. Econometrica 78, 883–931.

Deaton, Angus, Cartwright, Nancy, 2018. Reflections on randomized control trials. Soc. Sci. Med. 210, 86–90.

Devlin, B., Daniels, Michael, Roeder, Kathryn, 1997. The heritability of IQ. Nature 388, 468–471.

Dumais, Susan A., 2002. Cultural capital, gender, and school success: the role of *habitus*. Sociol. Educ. 75, 44–68.

Dumais, Susan A., Kessinger, Richard J., Ghosh, Bonny, 2012. Concerted cultivation and teachers' evaluations of students: exploring the intersection of race and parents' educational attainment. Sociol. Perspect. 55, 17–42.

Duncan, Greg J., Brooks-Gunn, Jeanne, 1997. Income effects across the life span: integration and interpretation. In: Duncan, G.J., Brooks-Gunn, J. (Eds.), Consequences of Growing up Poor. Russell Sage Foundation, New York, NY, pp. 596–610.

Duncan, Greg J., Magnuson, Katherine A., 2011. The nature and impact of early achievement skills, attention skills, and behavior problems. Pp. 47–70. In: Duncan, G. J., Murnane, R.J. (Eds.), Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances. Russell Sage Foundation, New York, NY.

Duncan, Greg J., Jean Yeung, W., Brooks-Gunn, Jeanne, Smith, Judith R., 1998. How much does childhood poverty affect the life chances of children? Am. Sociol. Rev. 63, 406–423.

Elder Jr., Glen H., Conger, Rand D., Michael, Foster E., Monika, Ardelt, 1992. Families under economic pressure. J. Fam. Issues 13, 5–37.

Michael, Gaddis S., 2013. The influence of *habitus* in the relationship between cultural capital and academic achievement. Soc. Sci. Res. 42, 1–13.

Gershoff, Elizabeth T., Lawrence Aber, J., Cybele Raver, C., Lennon, Mary Clare, 2007. Income is not enough: incorporating material hardship into models of income associations with parenting and child development. Child Dev. 78, 78–95.

Goldberg, Lewis R., Sweeney, Dennis, Merenda, Peter F., Edward Hughes Jr., John, 1998. Demographic variables and personality: the effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. Personal. Individ. Differ. 24, 393–403.

Guo, Guang, Harris, Kathleen Mullan, 2000. The mechanisms mediating the effects of poverty on children's intellectual development. Demography 37, 431–447.

Hauser, Robert M., Tsai, Shu-Ling, Sewell, William H., 1983. A model of stratification with response error in social and psychological variables. Sociol. Educ. 56, 20–46.

Heckman, James J., Stixrud, Jora, Urzúa, Sergio, 2006a. The effects of cognitive and noncognitive Abilities on labor market outcomes and social behavior. J. Labor Econ. 24, 411–482.

Heckman, James J., Urzua, Sergio, Vytlacil, Edward, 2006b. Understanding instrumental variables in models with essential heterogeneity. Rev. Econ. Stat. 88, 389–432.

Heckman, James J., Pinto, Rodrigo, Savelyev, Peter, 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Am. Econ. Rev. 103, 2052–2086.

Heckman James, J., 2006. Skill formation and the economics of investing in disadvantaged children. Science 312, 1900–1902.

Hsin, Amy, Xie, Yu, 2017. Life-Course changes in the mediation of cognitive and non-cognitive skills for parental effects on children's academic achievement. Soc. Sci. Res. 63, 150–165.
Irwin, Sarah, Sharon, Elley, 2011. Concerted cultivation? Parenting values, education and class diversity. Sociology 45, 480–495.
Kaushal, Neeraj, Magnuson, Katherine, Waldfogel, Jane, 2011. How is family income related to investments in children's learning. Pp. 187–205. In: Duncan, G.J., Murnane, R.J. (Eds.), Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances. Russell Sage Foundation, New York, NY.
Kautz, Tim, Heckman, James J., Diris, Ron, Weel, Bas ter, Borghans, Lex, 2014. Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success. National Bureau of Economic Research (NBER). Working Paper 20749. http://www.nber.org/papers/w20749.
Knudsen, Eric I., Heckman, James J., Cameron, Judy L., Shonkoff, Jack P., 2006. Economic, neurobiological, and behavioral perspectives on building America's future workforce. Proc. Natl. Acad. Sci. 103, 10155–10162.
Kohn, Melvin L., 1959. Social class and parental values. Am. J. Sociol. 64, 337–351.
Kohn, Melvin L., Schooler, Carmi, 1969. Class, occupation, and Orientation. Am. Sociol. Rev. 34, 659–678.
Lareau, Annette, 2011. Unequal Childhoods: Class, Race, and Family Life. University of California Press, Berkeley, CA.
Jung-Sook, Lee, Bowen, Natasha K., 2006. Parent involvement, cultural capital, and the achievement gap among elementary school children. Am. Educ. Res. J. 43, 193–218.
Lewis, Oscar, 1966. The Culture of Poverty. W. H. Freeman, San Francisco, CA.
Li, Weilin, Farkas, George, Duncan, Greg, Burchinal, Margaret, Vandell, Deborah, 2013. Timing of high-quality child care and cognitive, language, and preacademic development. Dev. Psychol. 49, 1440–1451.
Linver, Miriam R., Brooks-Gunn, Jeanne, Kohen, Dafna E., 2002. Family processes as pathways from income to young children's development. Dev. Psychol. 38 (5), 719–734.
Ludwig, Jens, Phillips, Deborah A., 2007. The Benefits and Costs of Head Start. National Bureau of Economic Research (NBER). Working Paper 12973. http://www.nber.org/papers/w12973.
Lugo-Gil, Julieta, Tamis-LeMonda, Catherine S., 2008. Family resources and parenting quality: links to children's cognitive development across the first 3 Years. Child Dev. 79, 1065–1085.
Manski, Charles F., Garfinkel, Irving, 1992. Introduction, pp. 1–21. In: Manski, Charles F., Garfinkel, Irwin (Eds.), Evaluating Welfare and Training Programs. Harvard University Press, Cambridge, MA.
Mayer, Susan E., 1997. What Money Can't Buy: Family Income and Children's Life Chances. Harvard University Press, Cambridge, MA.
McLeod, Jane D., Shanahan, Michael J., 1993. Poverty, parenting, and children's mental health. Am. Sociol. Rev. 58, 351–266.
McLanahan, Sara.S., Sandefur, Gary, 1994. Growing up with a Single Parent. Harvard University Press, Cambridge, MA.
Pearl, Judea, 2009. Causality: Models, Reasoning, and Inference, second ed. Cambridge University Press, New York, NY.
Pilarz, Ros, Alejandra, Hill, Heather D., 2014. Unstable and multiple child care arrangements and young children's behavior. Early Child. Res. Q. 29, 471–483.
Puma, Michael, Bell, Stephen, Cook, Ronna, Heid, Camilla, Shapiro, Gary, Broene, Pam, Jenkins, Frank, Fletcher, Philip, Quinn, Liz, Friedman, Janet, Ciarico, Janet, Rohacek, Monica, Adams, Gina, Spier, Elizabeth, 2010. Head Start Impact Study, Final Report, Administration for Children and Families U.S. Department of Health and Human Services, Washington, DC.
Plomin, Robert, DeFries, John C., Fulker, David W., 2006. Nature and Nurture during Infancy and Early Childhood. Cambridge University Press, New York, NY.
Cybele, Raver C., Gershoff, Elizabeth T., Lawrence Aber, J., 2007. Testing equivalence mediating models of income, parenting, and school readiness for white, black, and hispanic children in a national sample. Child Dev. 78, 96–115.
Reardon, Sean, 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In: Murnane, R., Duncan, G. (Eds.), Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances. Russell Sage Foundation, New York, NY, pp. 91–116.
Redford, Jeremy, Johnson, Jennifer A., Honnold, Julie, 2009. Parenting practices, cultural capital and educational outcomes: the effects of concerted cultivation on academic achievement. Race, Gender Class 16, 25–44.
Roubinov, Danielle S., Boyce, William Thomas, 2017. Parenting and SES: relative values or enduring principles? Curr. Opin. Psychol. 15, 162–167.
Ruzek, Erik, Burchinal, Margaret, Farkas, George, Duncan, Greg, 2014. The quality of toddler child care and cognitive skills at 24 Months: propensity score analysis results from the ECLS-B. Early Child. Res. Q. 29, 12–21.
Schweinhart, Lawrence J., Montie, Jeanne, Xiang, Zongping, Barnett, William S., Belfield, Clive R., Nores, Milagros, 2005. Lifetime Effects: the High/Scope Perry Preschool Study through Age 40. Monographs of the High/Scope Educational Research Foundation, 14. High/Scope Press, Ypsilanti, MI.
Schneider, Daniel, Hastings, Orestes P., LaBriola, Joe, 2018. Income inequality and class divides in parental investments. Am. Sociol. Rev. 83, 475–507.
Schweinhart, Lawrence J., Barnes, Helen V., Weikart, David P., Barnett, William S., Epstein, Ann S., 1993. Significant Benefits: the High/Scope Perry Preschool Study through Age 27. Monographs of the High/Scope Educational Research Foundation, 10. High/Scope Press, Ypsilanti, MI.
Sewell, William H., Haller, Archibald O., Portes, Alejandro, 1969. The educational and early occupational attainment process. Am. Sociol. Rev. 34, 82–92.
Stuart, Elizabeth A., Cole, Stephen R., Bradshaw, Catherine P., Leaf, Philip J., 2011. The use of propensity scores to assess the generalizability of results from randomized trials. J. R. Stat. Soc. Ser. A 174 (2), 369–386.
Subramanian, S.V., Kim, Rockli, Christakis, Nicholas A., 2018. The "average" treatment effect: a construct ripe for retirement. A commentary on Deaton and Cartwright. Soc. Sci. Med. 210, 77–82.
Sui-Chu, Ho, Esther, Douglas Willms, J., 1996. Effects of parental involvement on eighth-grade Achievement. Sociol. Educ. 69, 126–141.
Turkheimer, Eric, Haley, Andreana, Waldron, Mary, D'Onofrio, Brian, Gottesman, Irving I., 2003. Socioeconomic status modifies heritability of IQ in young children. Psychol. Sci. 14 (6), 623–628.
van Eijck, Koen, Paul de Graaf, M., 2004. The big five at school: the impact of personality on educational attainment. Neth. J. Soc. Sci. 40, 24–40.
Vandell, Lowe, Deborah, 2004. Early child care: the known and the unknown. Merrill-Palmer Q. 50, 387–414.
VanderWeele, Tyler, 2015. Explanations in Causal Inference. Oxford University Press, New York, NY.
VanderWeele, Tyler, 2016. Mediation analysis: a practitioner's guide. Annu. Rev. Public Health 37, 2.1–2.16.
VanderWeele, Tyler J., Robins, James M., 2007. Four types of effect modification: a classification based on directed acyclic graphs. Epidemiology 18 (5), 561–568.
Xie, Yu, 2013. Population heterogeneity and causal inference. Proc. Natl. Acad. Sci. 110 (16), 6262–6268.
Xie, Yu, Brand, Jennie, Jann, Ben, 2012. Estimating heterogeneous treatment effects with observational data. Sociol. Methodol. 42, 314–347.
Jean, Yeung W., Linver, Miriam R., Brooks-Gunn, Jeanne, 2002. Money matters for young children's development: parental investment and family processes. Child Dev. 73 (6), 1861–1879.