# Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes

James Heckman, Lena Malofeeva, Rodrigo Pinto, and Peter Savelyev*

The University of Chicago

March 4, 2010

* James Heckman is the Henry Schultz Distinguished Service Professor of Economics at the University of Chicago; Professor of Science and Society, University College Dublin; and Cowles Foundation Research Professor at Yale University. Lena Malofeeva is Director of Evaluation at the First Things First External Evaluation, College of Education, University of Arizona. Rodrigo Pinto and Peter Savelyev are Ph.D. candidates in economics at the University of Chicago. A version of this paper was presented at a seminar at the HighScope Foundation, Ypsilanti, Michigan, December 2006; at a conference at the Minneapolis Federal Reserve, December 2007; at a National Poverty Center conference, Ann Arbor, Michigan, December 2007; at a conference at Castle Marbach, Germany, April 2008, sponsored by the Jacobs Foundation; at the Leibniz Network Conference on noncognitive skills in Mannheim, May 2008; at an Institute for Research on Poverty conference, Madison, Wisconsin, June 2008; at the Society for Research on Child Development, Denver, April 2009; at the Association for Research in Personality Conference, Evanston, July 2009; at the Public Policy & Economics Workshop at the Harris School of Public Policy, the University of Chicago, October 2009; and at the Cultivating Human Capital Conference, Chicago, December 2009. We thank participants at these meetings. We are grateful to Clancy Blair, Sarah Cattan, Angela Duckworth, Miriam Gensowski, Jeff Grogger, Devesh Raval, Brent Roberts, Tino Sanandaji, Larry Schweinhart, Sandra Waxman, Ben Williams, and Junjian Yi for helpful comments. We are grateful to Christopher Hansman, Kegon Tan Teng Kok, Min Ju Lee, Xiliang Lin and Yun Pei for their excellent research assistance. This research was supported by the Committee for Economic Development through grants from the Pew Charitable Trusts and the Partnership for America's Economic Success; the JB & MK Pritzker Family Foundation; Susan Thompson Buffett Foundation; Mr. Robert Dugger; and NICHD (R01HD043411). We thank the HighScope Foundation for supplying us with initial psychological measures used in this paper. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here. Supplementary materials may be retrieved from http://jenni.uchicago.edu/Perry/.

**Abstract**

The Perry Preschool program was a randomized social experiment with long run followup that supplemented the early environments of disadvantaged African American children. Evidence from the program is widely used in support of claims of the effectiveness of early childhood intervention programs. The Perry program has been shown to boost adult performance for both girls and boys. It has a rate of return above the historical return to equity. The program had little lasting impact on the IQs of its participants. Using newly discovered data on the noncognitive traits of treatments and controls collected measured a few years after the program finished, we find that experimentally-induced changes in noncognitive traits explain a sizable portion of later-life treatment effects for education, employment, earnings, and crime. This paper contributes to an emerging literature on the importance of noncognitive traits in explaining a variety of social outcomes.

Key words: cognitive traits, noncognitive traits, factor analysis, human capital, human development, early childhood intervention, social experiment, preschool, disadvantaged children, Perry Preschool program

JEL codes: I21, I28, I29, J13, J15, J16, J24, 015

James Heckman
University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
(773) 702-0634
jjh@uchicago.edu

Lena Malofeeva
University of Arizona
College of Education
1501 E. Speedway
Tucson, AZ 85721
(520) 621-4487
lenam@email.arizona.edu

Rodrigo Pinto
University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
(773) 702-9436
rodrig@uchicago.edu

Peter Savelyev
University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
(773) 256-6351
psavel@uchicago.edu

# Introduction

A large literature establishes that children's early environments have substantial impacts on their life outcomes (see, e.g, Shonkoff and Phillips, 2000, Knudsen et al., 2006, Cunha et al., 2006, and Reynolds et al., 2009). However, the precise channels through which early environments affect child development are just beginning to be understood. Many early childhood interventions have little or no effect on adult IQs, yet they have large economic returns. These interventions must operate through other channels (Borghans et al., 2008; Heckman, 2008).

This paper presents experimental evidence on the effect of a flagship early childhood intervention program on the noncognitive capacities of its participants. The Perry program is the earliest preschool intervention program evaluated by the method of random assignment. Data are collected on participants through their early 40s.[1] Its positive effects on numerous lifetime outcomes for disadvantaged children have greatly influenced policy discussions (see, e.g., Shonkoff and Phillips, 2000).[2] It has a statistically significant economic rate of return for both boys and girls in the range of 6–10% per year, which is above the historical return to equity[3] (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010a).

The Perry program initially boosted IQs. However, this effect faded within a few years after the end of the two-year program, with no statistically significant effect remaining for males, and only a borderline significant effect remaining for females. See Figure 1 for three measures of IQ, discussed in greater detail below and in our Web appendix, which display the same pattern of a surge and then a decline with age. Although the program had little lasting

[1]The formal name of the program is the "HighScope Perry Preschool program", which we abbreviate to the "Perry program".

[2]The small sample size of the Perry experiment has led some researchers to question the validity of its results (see, e.g., Hanushek and Lindseth, 2009; Herrnstein and Murray, 1994). Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b) show that many treatment effects of the program are statistically significant for both boys and girls and survive adjustment for multiple hypothesis testing.

[3]The post-World War II stock market rate of return to equity is 5.8% (DeLong and Magin, 2009).

effect on IQ, it raised scores on achievement tests[4] and on measures of noncognitive traits,[5] as can be seen in Figure 2. We show that the increments in noncognitive traits generated by the program produce increases in achievement tests and promote other beneficial behaviors.

This paper addresses three questions about the mechanisms that produced the Perry results. First, how do traits change in response to experimentally manipulated early childhood investments? Second, how do the experimentally-induced changes in traits translate into changes in life outcomes? Finally, how do treatment effects differ across quantiles of outcome distributions? Do those with initially higher levels of traits benefit the most or the least?

We use a version of factor analysis to answer the first two questions. Using multiple measures of traits, we estimate factor scores and study how the experiment shifted these scores. Condensing information in this fashion partly compensates for the small sample size of the Perry study and explicitly accounts for errors of measurement in the proxies for traits.

To address the first question, we estimate the effects of the experimental intervention on cognitive and noncognitive traits. We find that the program raises the noncognitive traits of both males and females. It also slightly raises the cognitive skills of females, but not that of males.

To address the second question, we estimate the effects of cognitive and noncognitive traits on later life outcomes. We find substantial gender differences in the estimated effects. Enhancements of a noncognitive trait related to personal behavior (self-control and control of aggressive tendencies) are especially powerful in reducing crime which is a major source of benefits from the program especially for males (Belfield et al., 2006; Heckman et al., 2010a). For females, the effects of the Perry program are more broadly distributed across cognitive and multiple noncognitive traits. The estimated effects on a broader array of traits

[4] It is important to distinguish achievement tests from IQ tests, since they are often confused. While IQ tests are designed to measure "pure" general intelligence, achievement tests by construction measure acquired traits and knowledge. Acquired traits and knowledge depend on noncognitive traits such as motivation and conscientiousness. As a result, achievement tests are partly determined by noncognitive traits. Borghans et al. (2010) show that personality traits explain about 50% of the variance in many achievement tests.

[5] See Section II for details about the specific measures of noncognitive traits.

for females explains the wider variety of affected outcomes for them, as documented by Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b).

To address the third question, we assume that treatment preserves ranks in the distributions of outcomes for treatment and control groups. We calculate differences in means between the quantiles of outcomes for treatment and control groups. Treatment effects generally vary across quantiles of the outcome distributions. The effects on achievement scores and wages are stronger for those who would have had higher achievement scores and wages even without treatment. However, the effect on crime reduction is stronger for those with a higher propensity to commit crime.

This paper contributes to an emerging literature in economics documenting the predictive power of noncognitive traits (see Bowles and Gintis, 1976, 2001; Borghans, Duckworth, Heckman, and ter Weel, 2008; Heckman, Stixrud, and Urzua, 2006; Segal, 2008a, 2009). We depart from previous analyses of this issue in four ways: (a) by using experimental data; (b) by using cleaner measures of cognition (IQ rather than achievement tests); (c) by performing decompositions of treatment effects into components due to increments in cognitive and noncognitive traits; and (d) by using a simple regression-based approach instead of more complicated maximum-likelihood methods.

Our methodology is general and can be applied to other experiments to uncover channels through which treatment operates. The methodology has desirable statistical properties and is easy to implement. It accounts for measurement error, which is pervasive in measures of human performance.[6] Our experimental demonstration of the powerful role of noncognitive traits in determining various social outcomes is in agreement with recent non-experimental evidence presented by Duckworth and Seligman (2005)[7] and by Segal (2008a, 2009)[8].

[6] See, e.g., the evidence in Cunha and Heckman (2008), or Cunha, Heckman, and Schennach (2010).

[7] Duckworth and Seligman use data on 140 eighth-grade magnet public school students to demonstrate that self-discipline is a major predictor of educational outcomes. For the specific group they study, self-discipline explains about twice as much variance as IQ does.

[8] Segal uses data from the National Educational Longitudinal Survey to show the malleability of misbehavior and the long-run economic returns to improvements in this trait. She presents statistically significant associations between teacher-evaluated misbehavior measures at grade 8 and later life outcomes such as educational attainment and earnings, which remain even after controlling for family background.

Our empirical results contribute to the empirical literature on the effects of early childhood programs. We use previously unanalyzed item-level data on the noncognitive traits of both treatment and control groups to uncover mechanisms through which early intervention programs promote economic and social success.[9]

This paper proceeds as follows. Section I describes the Perry experiment and its emphasis on improving the self-discipline and social skills of its participants. Section II describes the measures that we use to proxy the underlying cognitive and noncognitive traits. Section III presents our methodology. Section IV reports our empirical results. Section V concludes. Supplementary material is placed in the Web Appendix.[10]

# I The Perry Program: Design and Background

The Perry program was a preschool educational intervention targeted towards African American children of low IQ and socioeconomic status. The program was conducted during the early to mid-1960s in the district of the Perry elementary school, a public school in Ypsilanti, Michigan. Children began the program at age three and were enrolled for two years.[11]

The 123 program participants were randomized into treatment and control groups via a complex protocol.[12] The Perry sample consisted of 51 females (25 treatments and 26 controls) and 72 males (33 treatments and 39 controls). There was relatively little attrition: only 11 participants had left the study by the time of the interview at age 40.[13]

The Perry curriculum was based on the principle of *active participatory learning*, in which

[9] Previous studies of the Perry project analyze only aggregates of the noncognitive measures used in this paper (Berrueta-Clement et al., 1984; Reynolds et al., 2009; Schweinhart et al., 1993; Weikart, 1967; Weikart et al., 1978b), and do not decompose treatment effects into components due to cognitive and noncognitive traits, as we do in this paper.

[10] http://jenni.uchicago.edu/Perry/

[11] The first entry cohort was enrolled for only one year of the program, beginning at age four.

[12] Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b) describe the protocol and develop statistical procedures for testing treatment effects taking into account the peculiarities of the Perry randomization protocol. They cite a study by Weikart, Epstein, Schweinhart, and Bond (1978) that show that there was virtually 100% participation in the experiment by all parents and children offered treatment in the Perry catchment area, so there was not greater participation in the program by more motivated parents.

[13] Five controls and two treatment participants died; two controls and two treatment group participants were missing.

children and adults were seen as equal partners in the learning process. Children engaged with objects, people, events, and ideas. Children's abilities in planning, executing and evaluating tasks were fostered, as were their social skills, including cooperation with others and resolution of interpersonal conflicts. The curriculum was grounded in research on cognitive development by Jean Piaget (Piaget and Inhelder, 2000), the progressive educational philosophy of John Dewey (Dewey, 1997), and the socio-cultural theories of Lev Vygotsky (Vygotsky, 1986).

Classes lasted 2.5 hours a day, five days a week during the school year. Teachers also made a weekly 1.5 hour home visit to each mother and child with the aim of involving the mother in the educational process and implementing the curriculum in the home. Web Appendix A provides a detailed description of the curriculum.

Data on both treatment and control participants were collected annually from age 3 to age 15. Three additional follow-ups were conducted when participants were 19, 27, and 40. Over the span of the study, numerous measures were collected on a variety of socioeconomic outcomes. IQ tests were given yearly at ages 3–9.[14] The noncognitive measures used in this paper were collected at ages 7–9.

Table 1 presents a comprehensive summary of background variables, lifetime outcomes and measures of traits used in this paper. Substantial differences between treatment and control groups occur for all categories of variables: for noncognitive measures (Personal Behavior and Socio-Emotional State, discussed further below); for outcomes (education, employment, crime, and other outcomes); and for two background variables (father's presence in the household and mother's employment). Differences between treatments and controls on mother's employment and father's presence emerged because of compromises in the randomization protocol. We account for these differences by controlling for background variables.[15]

[14]The Stanford-Binet test was given at ages 3–10.

[15]We use the methods of Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b), who analyze the treatment effects for the Perry data.

# II Our Measures of Cognitive and Noncognitive Traits

Human capacities are legion and, in general, they cannot be reduced to a scalar human capital. A large literature establishes the importance of cognition (measured by IQ) in explaining a variety of life outcomes (see, e.g., Jensen, 1998, and Gottfredson, 1997, for surveys). Borghans et al. (2008) survey the emerging literature on noncognitive traits. The term "noncognitive traits," traditional in the economics literature, is general and literally means "all traits excluding cognition." The Big Five personality framework (John and Srivastava, 1999; McCrae and John, 1992) has emerged as one schema within which to measure noncognitive traits. The Big Five was codified after the Perry experiment was conducted, and thus we are forced to use pre-Big Five psychological measures. Like Segal (2008b, 2009), we use teacher's evaluations of student behaviors. We also use teacher's evaluations of socio-emotional state of the students. These measures are related to Big Five Conscientiousness, Agreeableness and Neuroticism. but they are not exactly the Big Five measures as they have come to be known in modern personality psychology.

## II.A Our Measure of Cognition

We use the Stanford-Binet Intelligence Scale (Terman and Merrill, 1960) to proxy the cognition of participants.[16] Mean differences by age between treatments and controls in Stanford-Binet Scores are plotted in Figure 1, panels (a) and (b). A boost in the IQs of the children in the treatment group is observed soon after the program starts at age 3. After the end of the two-year program, the treatment effect on IQ essentially disappears for males. A small positive effect remains for females. In our analysis, we use Stanford-Binet IQ measured at ages 7–9, the period when the treatment effect on IQ became relatively stable for both genders.[17]

[16]The Stanford-Binet test is widely used and has high reliability (see Santrock, 2008). See Appendix B for a detailed description of the test.

[17]Alternative, less widely-used measures of IQ, like Leiter and PPVT, evolve in a fashion similar to the Stanford-Binet IQ (see panels (c-f) of Figure 1). We choose Stanford-Binet IQs as measures since they cover

## II.B Measures of Noncognitive Traits

We use two sets of noncognitive measures to proxy traits designated as "Personal Behavior" and "Socio-Emotional State" in this paper. These measures are drawn from the Pupil Behavior Inventory (PBI) (Vinter et al., 1966; Weikart et al., 1978a). The PBI questionnaire consists of multiple questions, called *items*, asked of teachers about students in their classes.[18] The items are specifically designed by psychologists to proxy personality traits. For example, one set of items asks about different ways students can misbehave, such as by stealing, cheating, and swearing. The answers to this set of items proxy a latent trait called "Personal Behavior." In our empirical work, we use two sets of PBI items as measures of Personal Behavior and Socio-Emotional State traits. Web Appendix C presents a detailed discussion of all of the PBI items.

We base our choice of the two noncognitive traits used in this paper on both theoretical and empirical considerations. The recent literature in economics has shown that Personal Behavior (often called "misbehavior") is both malleable and predictive of life outcomes (Segal, 2008a, 2009), which makes it a likely source of treatment effects. The literature in psychology shows that Personal Behavior contributes to academic achievement (see, e.g., Egeland et al., 1990; Jimerson et al., 1999, 2002; Richman et al., 1982). We also use a second set of PBI items "socioemotional state" related to the Big Five trait, Neuroticism,[19] which has been shown to be predictive of a variety of life outcomes including job satisfaction,

the highest number of participants and they are the most established measures. We discuss these alternative measures in Web Appendix B, and relate them to the Stanford-Binet test.

[18]For example, one of the PBI items is the following question to the teacher: "How often is the student lying or cheating?", with a multiple choice answer including (1) very frequently, (2) frequently, (3) sometimes, (4) infrequently, and (5) very infrequently.

[19]Big Five Neuroticism is defined as a trait that contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad and tense (John and Srivastava, 1999). Neuroticism measures the degree to which a person experiences the world as threatening and beyond their control (Hogan and Hogan, 2007). Illustrative adjectives for neuroticism are moody, touchy, nervous, self-doubting as opposed to unnervous, relaxed, calm, stable, confident and effective (Harvey et al., 1995). Items defining Socio-Emotional State should be correlated with Neuroticism, although we have no direct statistical evidence on this question, since Perry data has no measures of Big Five Neuroticism. The items that we use include the following PBI measures: "appears depressed," "appears generally happy," "friendly and well-received by other pupils" and "withdrawn and uncommunicative" (see Table 2).

career success, relationship commitment between partners, marriage, health and health-related behaviors (Donnellan et al., 2004; Judge et al., 1999; Kurdek, 1997; Monden and Kraykamp, 2006; Seibert and Kraimer, 2001; Vollrath et al., 1999). Experimentally-induced enhancements in "Personal Behavior" and "Socio-Emotional State" are expected to generate beneficial life outcomes.

We have access to other measures of noncognitive traits. However, our empirical model becomes computationally unstable when we use more than two personality factors (in addition to a cognitive factor) to explain program treatment effects. We attribute this instability to the small sample size of the Perry experiment. Web Appendix E uses statistical model selection criteria to establish that a model using Cognition, Personal Behavior and Socio-Emotional State is the best model for predicting the mean treatment effects among all possible three-factor models formed using the sets of alternative personality questions that we have at our disposal.[20]

We derive all of the noncognitive measures used in this paper from the raw data at ages 7–9.[21] Table 2 describes items that constitute the measurement system for the cognitive and noncognitive traits analyzed in this paper. It shows 3–4 closely related items, which act as proxies for each trait. Table 3 presents measures of association among the items within each scale.[22] The table shows that items in each scale used in the paper are indeed suitable

[20]The mean treatment effects used in this analysis are for all of the outcomes displayed in Figures 4 and 5, which are extensively discussed below. These are chosen because they are statistically significant.

[21]The longitudinal structure of the Perry experiment allows us to obtain within-sample information necessary to impute missing data on measures. Participants who were not tested or evaluated at a particular age were often tested or evaluated at ages close to the missing one. Assuming the stability of these measures over the period between ages 7 and 9, we form average scores for each person using the non-missing observations at ages 7, 8 and 9. We impute 10 additional PBI observations instead of dropping persons with at least one missing observation over years 7–9 from the sample. Our averaging procedure not only keeps the sample size as large as possible, but also reduces the noisiness of the measures.

[22]This table reports Cronbach alphas. Reporting Cronbach alphas (see Cronbach, 1951) for each scale is standard in the psychological literature, but is usually not done by economists. Cronbach $\alpha$ is a statistic that captures how well a set of items $\{M_i\}_{i=1}^{N}$ measures a latent trait. Cronbach alpha is the lower bound of the internal consistency reliability of measures that are proxies for a trait. The internal consistency reliability is defined as a square of correlation between the measured scale $S = \sum M_i$ and the underlying trait $\theta$ (Allen and Yen, 2002). According to Table 3, the correlations between the traits and the scales (equal to square roots of alphas) range from 0.85 to 0.94 for males and from 0.89 to 0.96 for females. The Cronbach alphas can also be interpreted as a correlation between the observed scale and a hypothetical alternative scale, measuring the same trait and based on the same number of hypothetical alternative items (Nunnally and

proxies for a common trait. We use a modification of the original PBI Personal Behavior scale as our measure of Personal Behavior.

The original PBI "Personal Behavior" scale [23] was defined by 6 items: "lying or cheating," "steals," "swears or uses obscene words," "absences or truancies," "inappropriate personal appearance," and "poor personal hygiene." In the empirical analysis of this paper, we eliminate two items named "personal appearance" and "hygiene" from the Personal Behavior measure (see Table 2).[24] Our modification barely changes the empirical results obtained using the original measures, but contributes to the interpretability of the factor extracted[25]

The original PBI "Socio-Emotional State" scale was defined by five items. In this paper we use only four (see Table 2). We dropped the item called "isolated, few or no friends," since this item had an unusually high number of missing-values. By dropping this item we gained 12 additional observations (which is a big gain when the sample size is 123), while the Cronbach's alpha among the remaining components barely changed.

# III Model Specification

## III.A The Model for Outcomes

Let $D$ denote treatment status: $D = 1$ if a person is in the treatment group, and $D = 0$ otherwise. Let random vector $\boldsymbol{\theta}_D = (\theta_D^C, \theta_D^B, \theta_D^S)'$ be a vector of traits: Cognition ($C$), Personal Behavior ($B$) and Socio-Emotional State ($S$), which may be enhanced by treatment, denoted by $D$. As noted above, elements of this vector are permitted to be freely correlated. Traits affected by treatment may, in turn, affect outcomes $Y_{k,D}$, where $k \in \mathcal{K}$, and $\mathcal{K} =$

Bernstein, 1994). According to Table 3, the alphas (i.e., the correlations) range from 0.73 to 0.88 for males and from 0.79 to 0.91 for females.

[23]Here a "scale" means a set of items designed to proxy a trait. Elsewhere in the paper the term also means either an unweighted sum of proxies of a trait or a collection of items such as the Ypsilanti Rating Scale designed to proxy several traits.

[24]We retain the convention of labeling the capability measure we use by its original name "Personal Behavior" to stress a strong relationship of our traits to the original traits (Vinter et al., 1966).

[25]We thank Brent Roberts, a leading personality psychologist at the University of Illinois, for suggesting this grouping of the measures.

$\{1, \cdots, K\}$.

Our model for outcome $Y_{k,D}$ is

$$Y_{k,D} = \tau_{k,D} + \boldsymbol{\alpha}_k \boldsymbol{\theta}_D + \boldsymbol{\beta}_k \boldsymbol{X}_D + \epsilon_{k,D}, \tag{1}$$

where $\tau_{k,D}$ is an intercept term, $\boldsymbol{\alpha}_k$ is a three-dimensional vector of factor loadings, $\boldsymbol{X}_D$ is a vector of pre-program control variables, $\boldsymbol{\beta}_k$ is vector of outcome-specific regression coefficients, and $\epsilon_{k,D}$ is independent across $k$ for all $D$ and is independent of the $\boldsymbol{\theta}_D$ and $\boldsymbol{X}_D$. It has a mean of zero.[26]

## III.B The Measurement System

A variable is called a measure if it is a proxy for a trait. Let the index set for measures associated with trait $j$ be $\mathcal{M}^j$. We denote measures that capture trait $j$ by $M^j_{m^j,D}$, where $m^j \in \mathcal{M}^j$. Each trait $j \in \mathcal{J}$, where $\mathcal{J} = \{C, B, S\}$, may be associated with a different number of measures. To set the scale of the factors, we normalize the factor loading associated with the first measure of each factor to unity.[27] Thus, we write:

$$M^j_{1,D} = \nu^j_1 + \theta^j_D + \boldsymbol{\gamma}^j_{1,D} \boldsymbol{X}_D + \eta^j_{1,D}, \tag{2}$$

$$M^j_{m^j,D} = \nu^j_{m^j} + \varphi^j_{m^j,D} \theta^j_D + \boldsymbol{\gamma}^j_{m^j,D} \boldsymbol{X}_D + \eta^j_{m^j,D}; m^j \in \mathcal{M}^j \setminus \{1\}, \forall j \in \mathcal{J}, \tag{3}$$

where $\nu^j_{m^j}$ is an intercept term, $\boldsymbol{\gamma}^j_{m^j,D}$ is a measure-specific regression coefficient, $\varphi^j_{m^j,D}$ is a measure-specific factor loading, and $\eta^j_{m^j,D}$ is a mean zero error term independent of $\theta^j_D$ and $\boldsymbol{X}_D$ and of the other $\eta^j_{m^j,D}$ terms. We use three or more measurements per trait to secure identification of the model.[28] We follow Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b)

[26] Following the analysis of Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b), conditioning on covariates also corrects for compromises in randomization.

[27] The choice of the first measure is arbitrary and determines the interpretation of the estimated loadings. The choice of the first measure is a normalization that defines the scale of the factor, but does not affect the decompositions performed in this paper.

[28] See Web Appendix F for a discussion of the conditions under which the model is identified.

and condition on $\boldsymbol{X}$ to control for covariates and to adjust the estimates for compromises in the randomization protocol.

## III.C Decomposing Treatment Effects into Contributions from Mediating Variables

We decompose the mean of the conditional treatment effect for outcome $k$ adjusted for preprogram variables $\boldsymbol{X}$,

$$\Delta_k \mid \boldsymbol{X} \equiv E(Y_{k,1} - Y_{k,0}) - \boldsymbol{\beta}_k E(\boldsymbol{X}_1 - \boldsymbol{X}_0), \tag{4}$$

into contributions due to cognitive, noncognitive and other factors that we cannot capture with our measurements:

$$\Delta_k \mid \boldsymbol{X} = \underbrace{\alpha_k^C E(\theta_1^C - \theta_0^C)}_{\text{Cognition}} + \underbrace{\alpha_k^B E(\theta_1^B - \theta_0^B)}_{\text{Personal Behavior}} + \underbrace{\alpha_k^S E(\theta_1^S - \theta_0^S)}_{\text{Socio-Emotional State}} + \underbrace{(\tau_{k,1} - \tau_{k,0})}_{\text{Other Factors}}. \tag{5}$$

This decomposition is invariant with respect to any linear transformation of the measures $M^j_{m^j,D}$.[29]

## III.D Estimation Procedure

We estimate the model using a three-step procedure: 1) Estimate the parameters of the measurement system (2)–(3); 2) Use the parameters estimated in step 1 to estimate factor scores by the Bartlett (1937) method;[30] 3) Use the factor scores as regressors in OLS regressions

[29] A property of a standard factor model with intercepts is that an expression $\boldsymbol{\alpha}_k\boldsymbol{\theta}$ is invariant with respect to any linear transformation of measures that proxy $\boldsymbol{\theta}$. Consequently, in a setting when factors differ by treatment status, an expression $\boldsymbol{\alpha}_k E(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$ is invariant as well, since any linear change in level is differenced out $(E(\boldsymbol{\theta}_1 + \boldsymbol{\delta} - (\boldsymbol{\theta}_0 + \boldsymbol{\delta})) = E(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0))$. The invariance with respect to a change in scale is analogous to the standard case.

[30] Appendix G discusses the Bartlett method that we use to estimate the factor scores. We prefer the Bartlett (1937) approach to its main alternative, the Thomson (1934) approach, because the latter is biased unlike the former (Anderson and Rubin, 1956, Gorsuch, 1983, and Croon, 2002). See also Bolck et al. (2004); Lu and Thomas (2008); Skrondal and Laake (2001).

explaining the outcomes $Y_k$ correcting for errors in measurement that arise from using an estimated factor instead of the true factor.[31] Our procedure does not impose any statistical independence assumption between $\boldsymbol{\theta}$ and $\boldsymbol{X}$. We bootstrap to obtain standard errors.

Outcomes of the first and the second steps of our estimation procedure are not only necessary ingredients for the third step, but are also important objects of interest in their own right. In this paper we test hypotheses about equality of means of latent traits between the treated and the controlled and compare the estimated distributions of traits by treatment status and by gender (see Figure 3 and the associated analysis in Section IV.A).

By separating the estimation of measurement equations from the estimation of outcome equations, we isolate two conceptually different effects: (1) The effect of the experiment on changing traits; and (2) The effect of the experimentally induced changes in traits on changes in outcomes. A joint estimation of measurement and outcome equations could make the derived factors dependent on both the outcomes and the measures, and thus could render any causal interpretation of estimated effects of changes in the factors on the outcomes suspect. In our three-step procedure, factor scores are estimated using only the measurement equations. We determine how the experiment changed these measures and the latent traits they proxy. Another advantage of the three-step procedure over a joint estimation procedure is its computational simplicity.

# IV Empirical Results

We present three distinct empirical analyses. First, we study how treatment affects the marginal distributions of cognitive and noncognitive traits. Second, we decompose mean treatment effects on outcomes into components corresponding to changes in each trait. Finally, we split distributions of outcomes into quantiles and compare treatment effects calculated for different quantiles.

The first analysis is based on the first two steps of our three-step estimation procedure

[31] Web Appendix G presents a derivation of our error-adjustment procedure.

described in Section III. The second analysis uses the third step of the procedure. The third analysis does not rely on results from the three-step procedure.

## IV.A The Effect of the Perry Program on Psychological Traits

The Perry program has statistically significant[32] effects on the means of two latent traits (Cognition ($C$) and Personal Behavior ($B$)) for females, while only Personal Behavior is significantly affected for males.[33] The Effect on Socio-Emotional State ($S$) is borderline statistically significant for both genders.[34] These results are shown in Figure 3, which presents kernel densities of factor scores and one-sided $p$-values for the test of true equality of means of the psychological traits between treatment and control groups.

For the Personal Behavior and Socio-Emotional State of females, the treatment effect on noncognitive factors acts primarily through the enhancement of very low levels of the traits. The effect is clearly visible in Figure 3, panels (d) and (f).[35] In contrast, Cognition of females increased mostly for those who already had average and higher IQs[36] (see panel (b)). Panels (c) and (e) are consistent with boosting all levels of traits: low, average, and high. Thus, the patterns of treatment effects differ by trait and gender.

## IV.B Decomposing Treatment Effects on Outcomes by the Source

Tables 4–6 report the contributions of each of the innovations in Cognition, Personal Behavior, Socio-Emotional State, and other factors to the explanation of total treatment effects. Figures 4–6 are simplified visualizations of the tables, with the total treatment effect nor-

[32]Statistically significant at 10% level for Cognition of females, at 5% level for female Personal Behavior, and at 1% level for male Personal Behavior.

[33]Recall that we use measures at ages 7–9 to proxy noncognitive traits. For these ages the treatment effect on IQ can be considered as stable. In addition, for these ages both cognitive and noncognitive measures are available.

[34]The $p$-value is 0.151 for females and 0.177 for males.

[35]Some ceiling effect might be present, but elimination of the lower scores is clear.

[36]These IQs are average and high for a selected group of low-IQ children.

malized to 100%.[37]

Figure 4 and Table 4 show that for males, the intervention operates primarily through its effect on Personal Behavior. Contributions of Personal Behavior are generally statistically significant or borderline statistically significant (see Table 4), and explain up to 74% of the total treatment effects on lifetime outcomes (see Figure 4). Most of the statistically significant lifetime treatment effects for males are related to crime. Since changes in Personal Behavior are strongly predictive of changes in crime, our findings explain why crime reduction is a major benefit of the program for males (Belfield et al., 2006; Heckman et al., 2010a).

While changes in only one psychological trait explain the treatment effects for males, changes in all three traits explain the treatment effects for females, with their roles differing by outcome (see Figure 5 and Table 5).[38] For instance, for females, 26–53% of the achievement test treatment effect is explained by boosts in Cognition. The effect is statistically significant at age 7 and borderline statistically significant at ages 8 and 14 (see Table 5). Effects of Cognition on employment at age 19 and on "ever being on welfare" are borderline statistically significant.

Experimentally induced changes in Personal Behavior are important determinants of the treatment effects of marriage and crime outcomes for females. Socio-Emotional State has borderline statistically significant effects on "high school graduation" and on "being jobless." The effect of Socio-Emotional State on education and crime outcomes is not precisely determined (see Table 5).

It is instructive to distinguish two types of gender differences in the estimates from the Perry study. First, the sets of outcomes that exhibit at least borderline statistically significant treatment effects differ greatly across genders (see also Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010b). Second, even for outcomes showing at least borderline statistically sig-

[37]In the figures, contributions opposite to the sign of the total treatment effect are set to zero. Such contributions are small and statistically insignificant, so the figures present an accurate summary of the essential information in the tables. Web Appendix I discusses in depth the methodology for constructing the figures. See also the goodness of approximation displayed in Figure I-2 of the Web Appendix.

[38]We observe statistically significant effects at 10% level for Cognition and Personal Behavior of females. We observe borderline statistically significant effects for Socio-Emotional State.

nificant treatment effects for both genders, the effects due to cognitive and noncognitive components differ.

The first type of gender difference can be observed by comparing Figures 4 and 5. In those figures different sets of outcomes are decomposed by source, and different psychological traits dominate the decompositions, as described above. The source of this type of gender difference is the different malleabilities of psychological traits for males and females. For females, all three traits are affected, which results in a wide variety of affected outcomes. For males, only Personal Behavior is affected, resulting in a smaller variety of affected outcomes.

The second type of difference can be observed in Figure 6 and in Table 6, comparing decompositions by gender for the same set of outcomes. The figure shows that differences in treatment effects between genders also arise from gender differences in the role of Cognition and Socio-Emotional State, which explain some of the experimental outcomes for females, but none for males.[39] Changes in Personal Behavior affect the criminal activity of males and females in a comparable fashion.

## IV.C Quantile Treatment Effects

Quantile treatment effects assume preservation of ranks across potential outcome distributions.[40] Our analysis of the quantile treatment effects of the program shows that the program has generally stronger effects for those with higher achievement and wages, and for those more involved in crime.

Figure 7 shows unconditional differences in means for the whole sample (denoted by "total"), the 2-quantiles (denoted "lower 1/2" and "upper 1/2"), and the terciles (denoted "lower 1/3", "middle 1/3" and "upper 1/3"). For interpretative convenience, the higher quantile is defined to correspond to a more socially-desirable performance (higher wage,

[39] From Figure 6 one might falsely infer that Cognition explains CAT for males. In fact, the effect due to a boost in cognition is statistically insignificant, as documented by Table 6.

[40] Thus, when the program is applied to a group of people that can be ranked with respect to certain outcome like an achievement test, the first subject remains the first, the last remains the last, the median remains the median and the like across both treatment and control distributions. See Abbring and Heckman (2007) for a discussion of this widely-used rank-invariance assumption and its limitations.

lower crime, and so on).

Treatment effects on achievement tests and wages tend to be stronger for the upper quantile groups for males, i.e., those with higher levels of performance on the outcome (see panels (a) and (c) of Figure 7). For females this effect is much weaker, if present at all (see panels (b) and (d) of the same figure). For both genders, treatment effects on crime outcomes (panels (e) and (f)) are stronger for the lower quantile groups (those who commit more crimes whether treated or not).[41]

# V Conclusions

Economists often assume that human capital is a one-dimensional characteristic, proxying it with a measure such as the number of years of education or a test score (see Hanushek and Woessmann, 2008; Todd and Wolpin, 2003, 2007). Many scholars treat IQ as if it is the only measure of human potential (see, for instance, Herrnstein and Murray, 1994; Jensen, 1998). Education and IQ are indeed powerful predictors of success in life, but the assumption of a single dimension of human capital is greatly at odds with the evidence (see the survey by Borghans et al., 2008). This paper shows that changes in noncognitive traits resulting from the Perry program explain a substantial portion of its treatment effects. Our analysis reveals the multiplicity of traits that produce human achievement.

This research contributes to a large literature on early childhood development (see Shonkoff and Phillips, 2000, and Cunha et al., 2006). Using data from the Perry program, the most influential early childhood intervention evaluated by the method of random assignment with long-term follow-up, we analyze the mechanisms through which early investments produce substantial lifetime effects. We consider three traits: Cognition, Personal Behavior, and Socio-Emotional State. Our analysis decomposes Perry treatment effects on life outcomes

[41]For crime outcomes, the sample has a large number of zeros. Therefore, we observe little or no treatment effect for upper percentile groups. As a result of the decreased variability in the upper percentile groups, we obtain small standard errors. We are not able to provide a reliable inference for estimates in cases with reduced variability like the upper quantiles of crime. Hence, the reported standard errors in these cases should be interpreted with caution.

into components attributable to increments in these traits.

We offer a new perspective on the mechanism behind the surprising fact that a few hours per day of preschool at ages 3 and 4 with a curriculum that promotes social skills, planning, and organization, can significantly and beneficially affect a variety of lifetime outcomes. Noncognitive traits are enhanced by the intervention and play a substantial role in generating Perry's successful outcomes. The importance and malleability of these traits deserve much greater emphasis in public policy discussions.

# References


Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5145–5303. Amsterdam: Elsevier.

Allen, M. J. and W. M. Yen (2002). *Introduction to Measurement Theory.* Prospect Heights, IL: Waveland Press.

Anderson, T. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 5, pp. 111–150. Berkeley: University of California Press.

Bartlett, M. S. (1937, July). The statistical conception of mental factors. *British Journal of Psychology 28*(1), 97–104.

Belfield, C. R., M. Nores, W. S. Barnett, and L. Schweinhart (2006). The High/Scope Perry Preschool program: Cost-benefit analysis using data from the age-40 followup. *Journal of Human Resources 41*(1), 162–190.

Berrueta-Clement, J. R., L. J. Schweinhart, W. S. Barnett, A. S. Epstein, and D. P. Weikart

(1984). *Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19*. Ypsilanti, MI: High/Scope Press.

Bolck, A., M. Croon, and J. Hagenaars (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis 12*(1), 3–27.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2010). IQ, achievement, and personality. Unpublished manuscript, University of Maastricht and University of Chicago (revised from the 2009 version).

Bowles, S. and H. Gintis (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. New York: Basic Books.

Bowles, S. and H. Gintis (2001). Inheritance of economic status: Education, class, and genetics. In N. J. Smelser and P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*, pp. 4132–4141. Oxford, UK: Pergamon.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika 16*(3), 297–334.

Croon, M. A. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides and I. Moustaki (Eds.), *Latent Variable and Latent Structure Models*, pp. 195–223. NJ: Lawrence Erlbaum Associates, Inc.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. Forthcoming, *Econometrica*.

DeLong, J. and K. Magin (2009, Winter). The U.S. equity return premium: Past, present and future. *Journal of Economic Perspectives 23*(1), 193208.

Dewey, J. (1997). *Experience and Education*. New York: Free Press.

Donnellan, M. B., R. D. Conger, and C. M. Bryant (2004). The big five and enduring marraiges. *Journal of Research in Personality 38*, 481–504.

Duckworth, A. L. and M. E. P. Seligman (2005, November). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science 16*(12), 939–944.

Egeland, B., M. Kalkoske, N. Gottesman, and M. F. Erickson (1990). Preschool behavior problems: Stability and factors accounting for change. *Journal of Child Psychology and Psychiatry 31*, 891–910.

Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gottfredson, L. S. (1997, January-February). Why g matters: The complexity of everyday life. *Intelligence 24*(1), 79–132.

Hanushek, E. and A. A. Lindseth (2009). *Schoolhouses, Courthouses, and Statehouses: Solving the Funding-Achievement Puzzle in America's Public Schools*. Princeton, NJ: Princeton University Press.

Hanushek, E. and L. Woessmann (2008, September). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–668.

Harvey, R. J., W. D. Murry, and S. E. Markham (1995). A "Big Five" scoring system for the Myers-Briggs type indicator. Presented at the Conference of the Society for Industrial and Organizational Psychology, May, 1995. Orlando, Florida.

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

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). A reanalysis of the HighScope Perry Preschool Program. First draft, September, 2006. Forthcoming, *Quantitative Economics*.

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

Herrnstein, R. J. and C. A. Murray (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hogan, R. and J. Hogan (2007). *Hogan Personality Inventory Manual* (3rd ed.). Tulsa, OK: Hogan Assessment Systems.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Jimerson, S., B. Egeland, and A. Teo (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology 91*(1), 116–126.

Jimerson, S. R., P. Ferguson, A. D. Whipple, G. E. Anderson, and M. J. Dalton (2002). Exploring the association between grade retention and dropout: A longitudinal study examining socio-emotional, behavioral, and achievement characteristics of retained students. *The California School Psychologist 7*, 51–62.

John, O. P. and S. Srivastava (1999). The big five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research*, Chapter 4, pp. 102–138. New York: The Guilford Press.

Judge, T. A., C. A. Higgins, C. J. Thoresen, and M. R. Barrick (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology 52*(3), 621–652.

Knudsen, E. I., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

Kurdek, L. A. (1997). Relation between neuroticism and dimensions of relationship commitment: Evidence from gay, lesbian, and heterosexual couples. *Journal of Family Psychology 11*(1), 109–124.

Lu, I. R. R. and D. R. Thomas (2008). Avoiding and correcting bias in score-based latent variable regression with discrete manifest items. *Structural Equation Modelling 15*, 462–490.

McCrae, R. R. and O. P. John (1992, June). An introduction to the five-factor model and its applications. *Jounral of Personality 60*(2), 175–215.

Monden, C. and G. Kraykamp (2006, September). Neuroticism, education and self-assessed health in the general population of the united states. can smoking behavior explain the associations? *Social Indicators Research 78*(2), 271–285.

Nunnally, J. C. and I. H. Bernstein (1994). *Psychometric Theory*. New York: McGraw-Hill.

Piaget, J. and B. Inhelder (2000). *The Psychology of the Child*. New York: Basic Books.

Reynolds, A. J., M. Englund, S.-R. Ou, L. J. Schweinhart, and F. Campbell (2009). Paths of effects of preschool to educational attainment. Unpublished manuscript, University of Minnesota and University of North Carolina.

Richman, N., J. E. Stevenson, and P. J. Graham (1982). *Preschool to school: A behavioural study.* San Diego: Academic Press.

Santrock, J. W. (2008). *A Topical Approach to Life-Span Development* (Fourth ed.). New York: McGraw-Hill.

Schweinhart, L. J., H. V. Barnes, and D. P. Weikart (1993). *Significant Benefits: The High-Scope Perry Preschool Study Through Age 27.* Ypsilanti, MI: High/Scope Press.

Segal, C. (2008a, Fall). Classroom behavior. *Journal of Human Resources 43*(4), 783–814.

Segal, C. (2008b, October). Motivation, test scores, and economic success. Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain.

Segal, C. (2009). Misbehavior, education, and labor market outcomes. Unpublished manuscript, Universitat Pompeu Fabra, Department of Economics and Business.

Seibert, S. E. and M. L. Kraimer (2001). The five-factor model of personality and career success. *Journal of Vocational Behavior 58*, 1–21.

Shonkoff, J. P. and D. Phillips (2000). *From Neurons to Neighborhoods: The Science of Early Child Development.* Washington, DC: National Academy Press.

Skrondal, A. and P. Laake (2001). Regression among factor scores. *Psychometrika 66*(4), 563–576.

Terman, L. M. and M. A. Merrill (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M.* Boston: Houghton Mifflin.

Thomson, G. H. (1934, May). Hotelling's Method modified to give Spearman's g. *Journal of Educational Psychology 25*(5), 366–374.

Todd, P. E. and K. I. Wolpin (2003, February). On the specification and estimation of the production function for cognitive achievement. *Economic Journal 113*(485), F3–33.

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

Vinter, R. D., R. C. Sarri, D. J. Vorwaller, and W. E. Shafer (1966). *Pupil Behavior Inventory: A Manual for Administration and Scoring.* Ann Arbor, MI: Campus Publishers.

Vollrath, M., D. Knoch, and L. Cassano (1999). Personality, risky health behaviour, and perceived susceptibility to health risks. *European Journal of Personality 13*, 39–50.

Vygotsky, L. S. (1986). *Thought and Language.* Cambridge, MA: MIT Press.

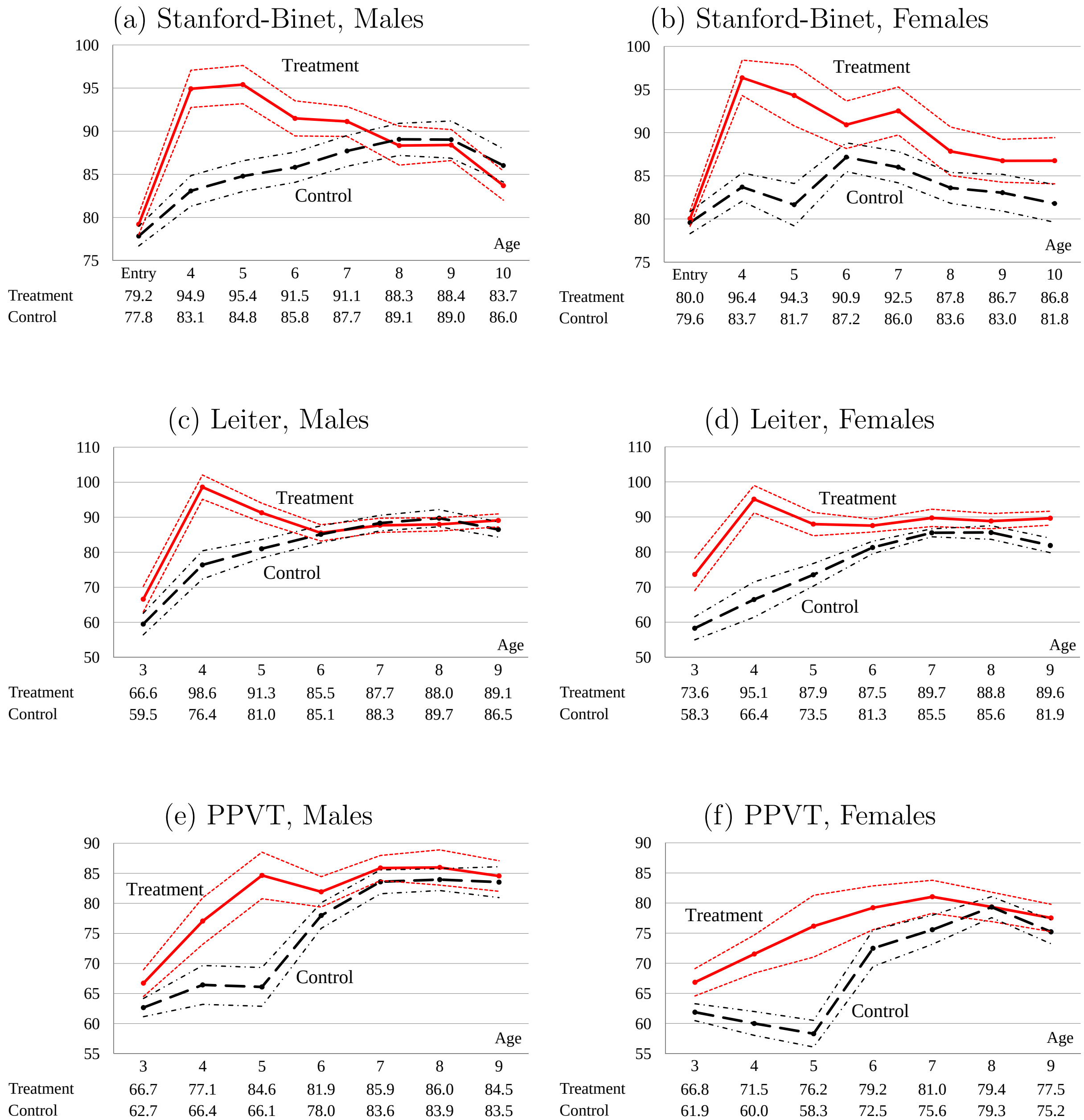
Weikart, D. P. (1967). Preliminary results from a longitudinal study of disadvantaged preschool children. Unpublished manuscript, ERIC No. ED 030 490. Presented at the 1967 convention of the Council for Exceptional Children, St. Louis, MO.

Weikart, D. P., J. T. Bond, and J. T. McNeil (1978a). *Monographs of the High/Scope Educational Research Foundation.* Ypsilanti, MI: The High/Scope Press.

Weikart, D. P., J. T. Bond, and J. T. McNeil (1978b). *The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results Through Fourth Grade.* Ypsilanti, MI: Monographs of the High/Scope Educational Research Foundation.

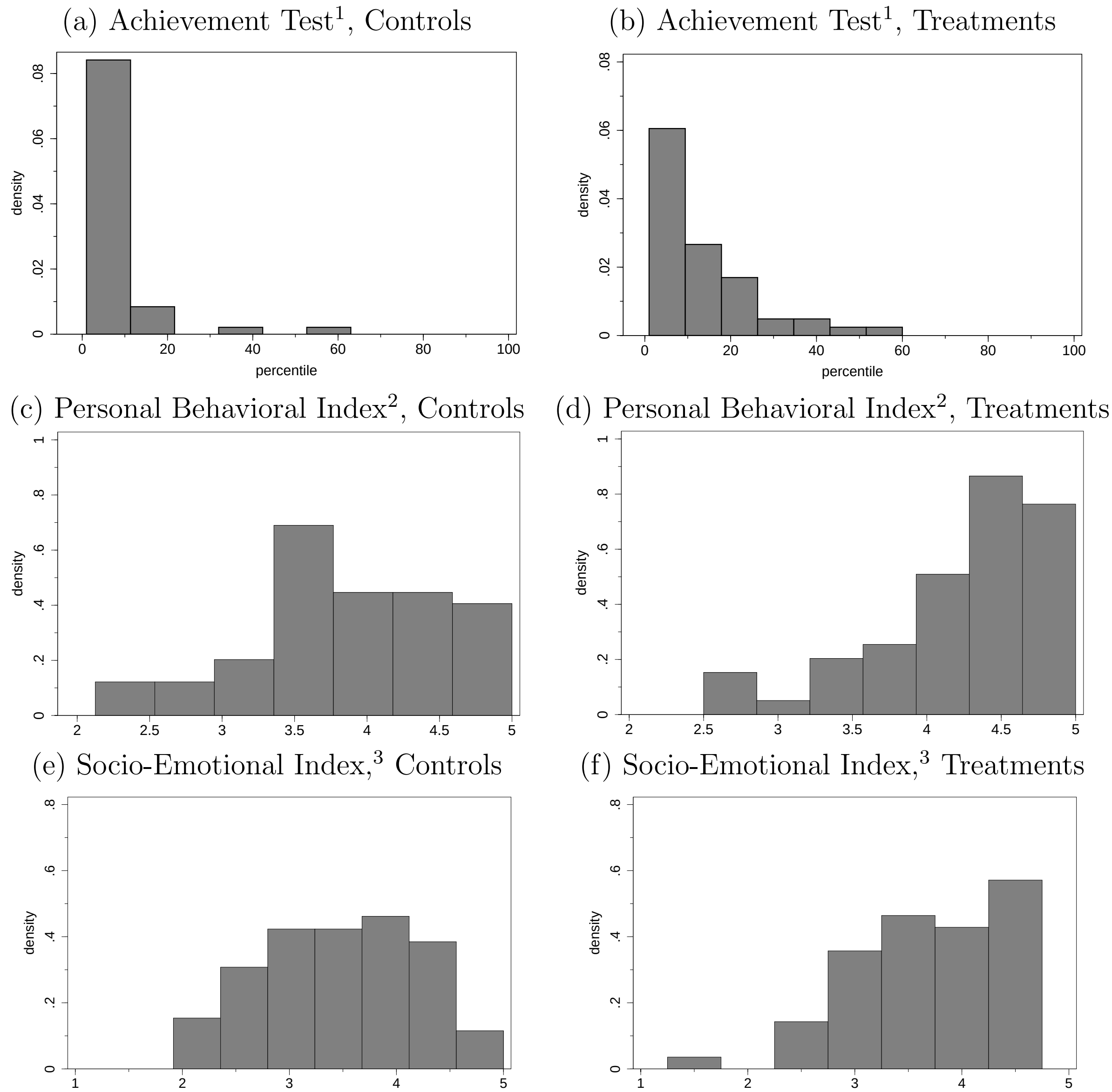
Weikart, D. P., A. S. Epstein, L. J. Schweinhart, and J. T. Bond (1978). *The Ypsilanti Preschool Curriculum Demonstration Project: Preschool Years and Longitudinal Results.* Ypsilanti, MI: High/Scope Press.

**Figure 1:** Cognitive Tests by Gender and Treatment Status



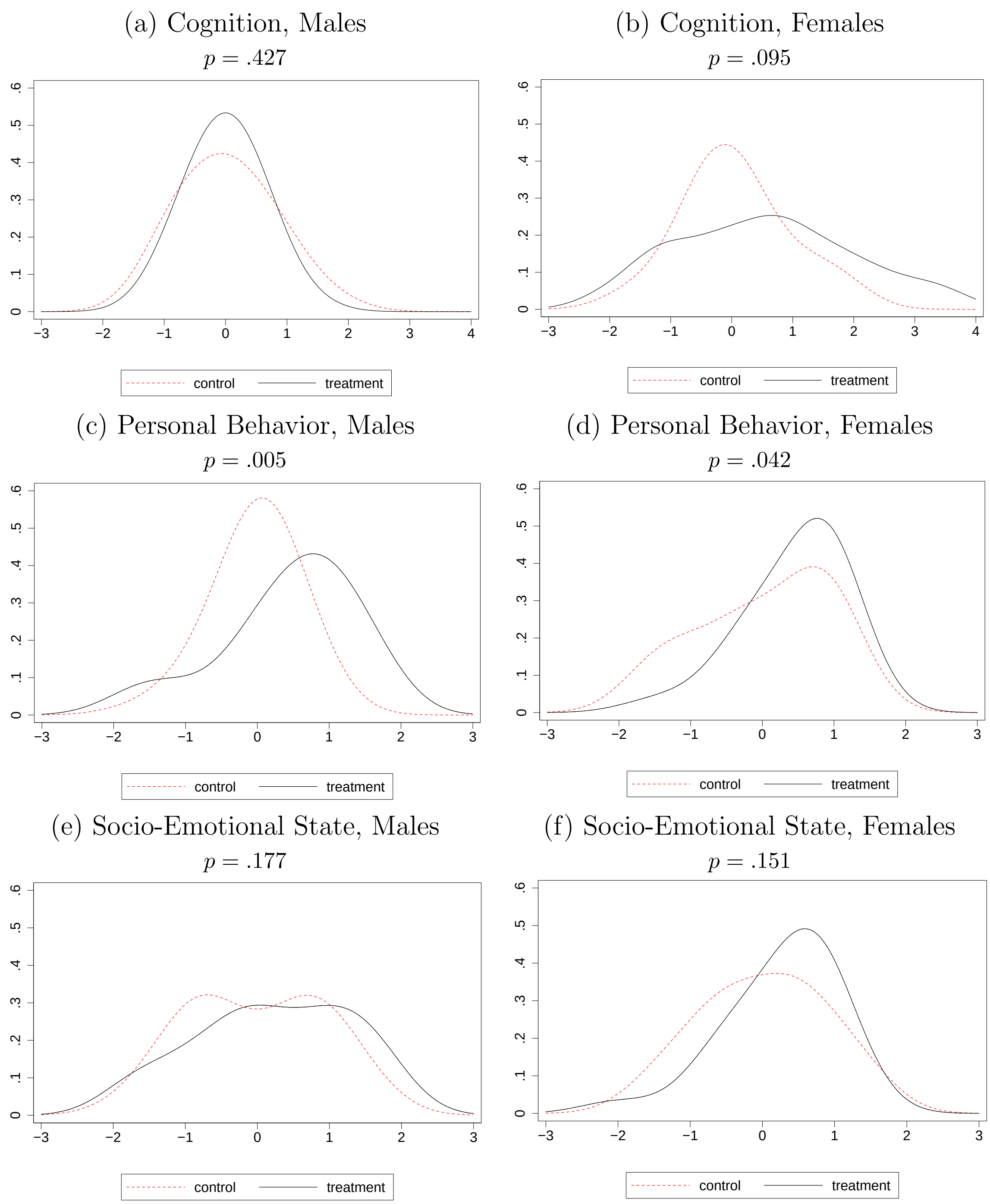

**Notes:** Stanford-Binet Intelligence Scale, Leiter International Performance Scale (Leiter), and Peabody Picture Vocabulary Test (PPVT) scores are shown for the Perry sample. Bold lines represent the means of IQs. Fine lines represent standard errors for the corresponding means: one standard error above and below the mean. For a detailed description of these cognitive measures, see Web Appendix B.

**Figure 2:** Histograms of Three Indicators



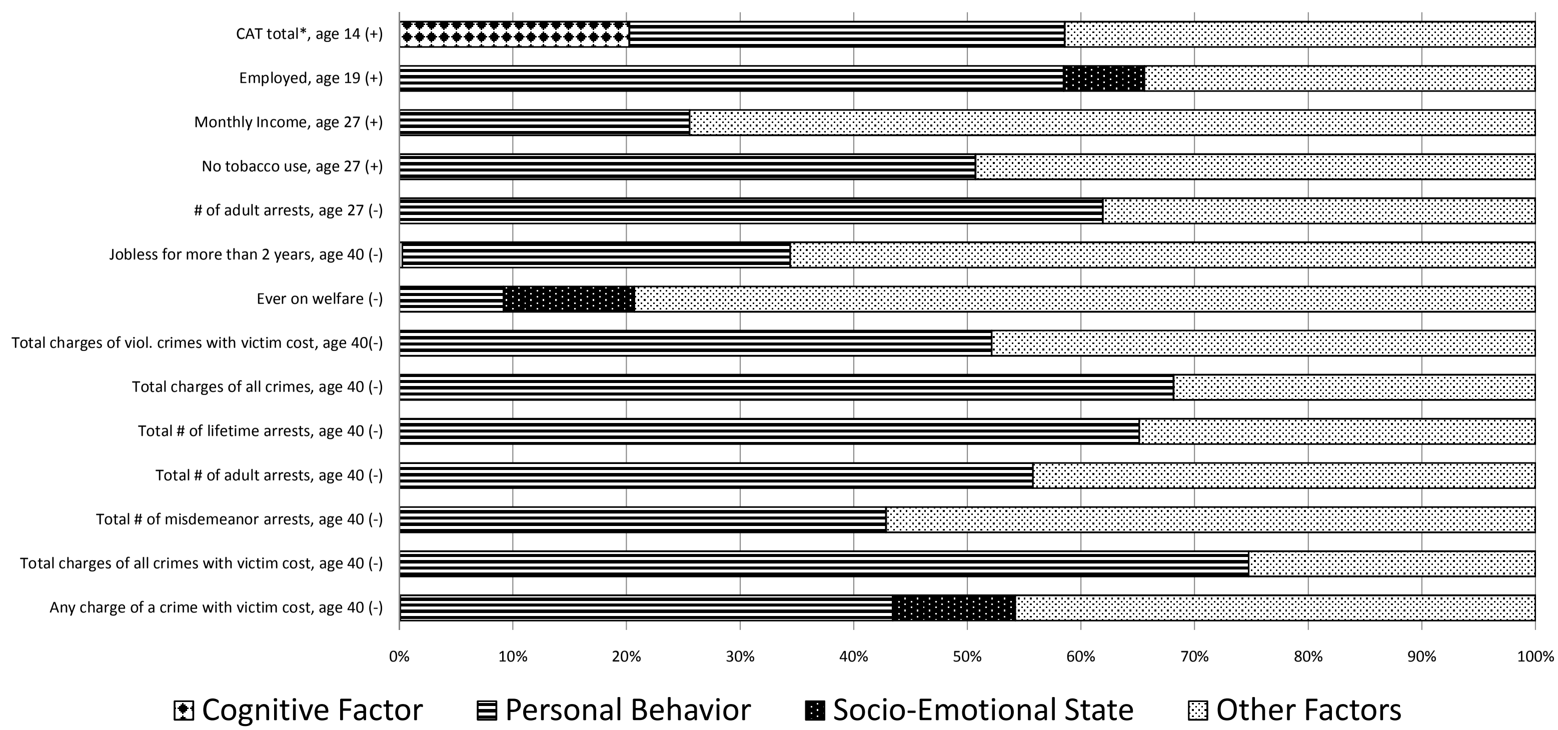

**Notes:** [1]Achievement test is the California Achievement Test (CAT) taken at age 14; [2]The Personal Behavior index is an unweighted average of four measures: "absences or truancies", "lying or cheating", "steals", and "swears or uses obscene words". Higher numbers correspond to a more socially-desirable behavior; [3]The Socio-Emotional index is an unweighted average of four measures: "appears depressed", "withdrawn and uncommunicative", "friendly and well-received by pupils", and "appears generally happy". Higher numbers correspond to a more socially-desirable behavior or trait. See Tables 1, 2 and 3 for more information about these indexes. The Web Appendix plots these histograms by gender in Figures B-1, C-5, and C-6. One-sided $p$-values for differences in means are 0.001, 0.002 and 0.096 for the Achievement test, Personal Behavior index, and Socio-Emotional index.

**Figure 3:** Kernel Densities of Factor Scores



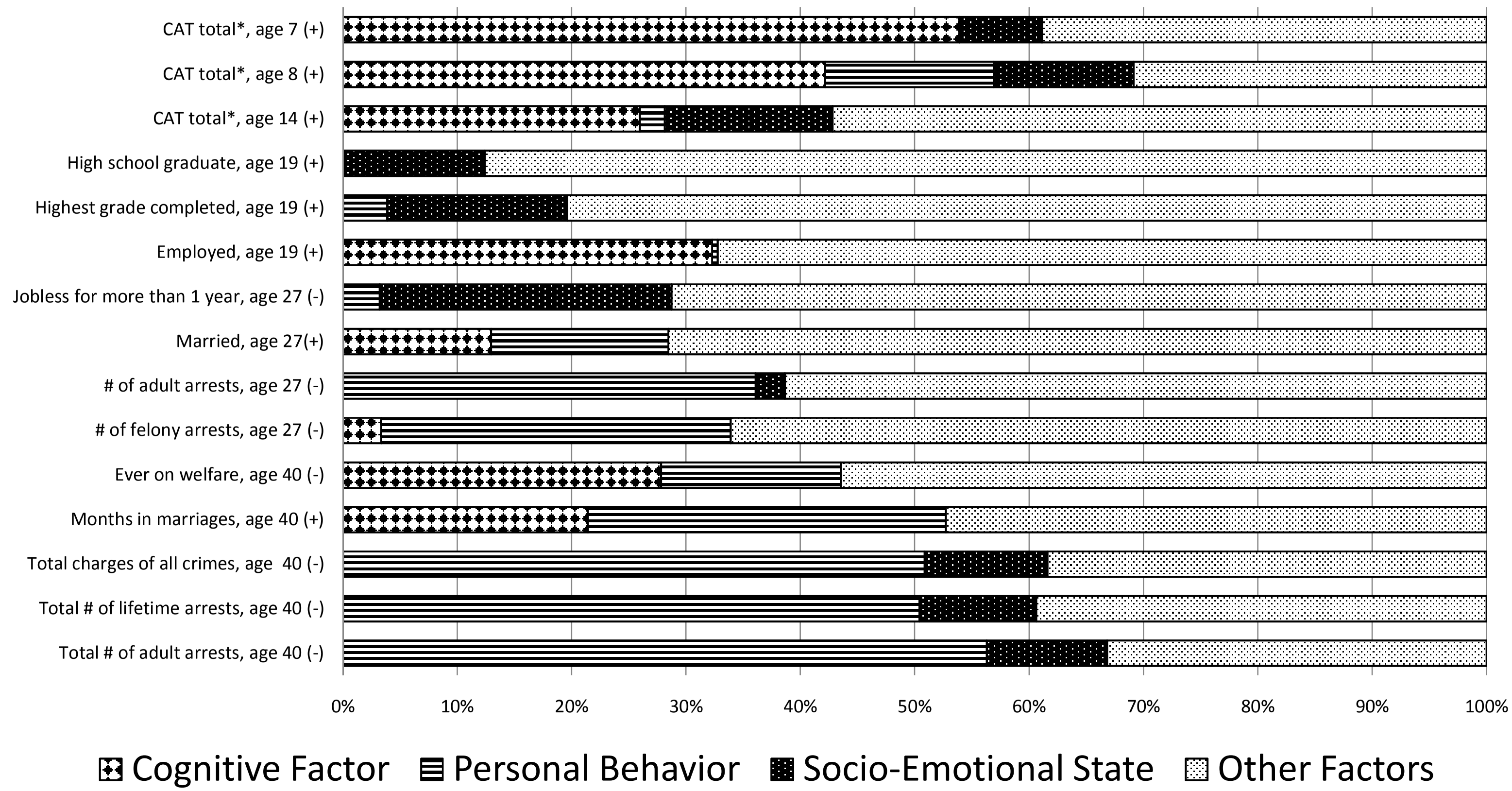

**Notes:** Numbers above the charts are one-sided bootstrap $p$-values testing the equality of factor score means for the treated and the controlled. Higher numbers correspond to more socially-desirable behaviors or traits. Kernel density functions of Bartlett (1937) factor scores, which are used for calculations, are shown. The number of bootstrap draws is 1000. Calculations are based on the Perry data. See Figure I-1 of the Web Appendix for the corresponding empirical CDFs.

**Figure 4:** Decompositions of Treatment Effects, Males



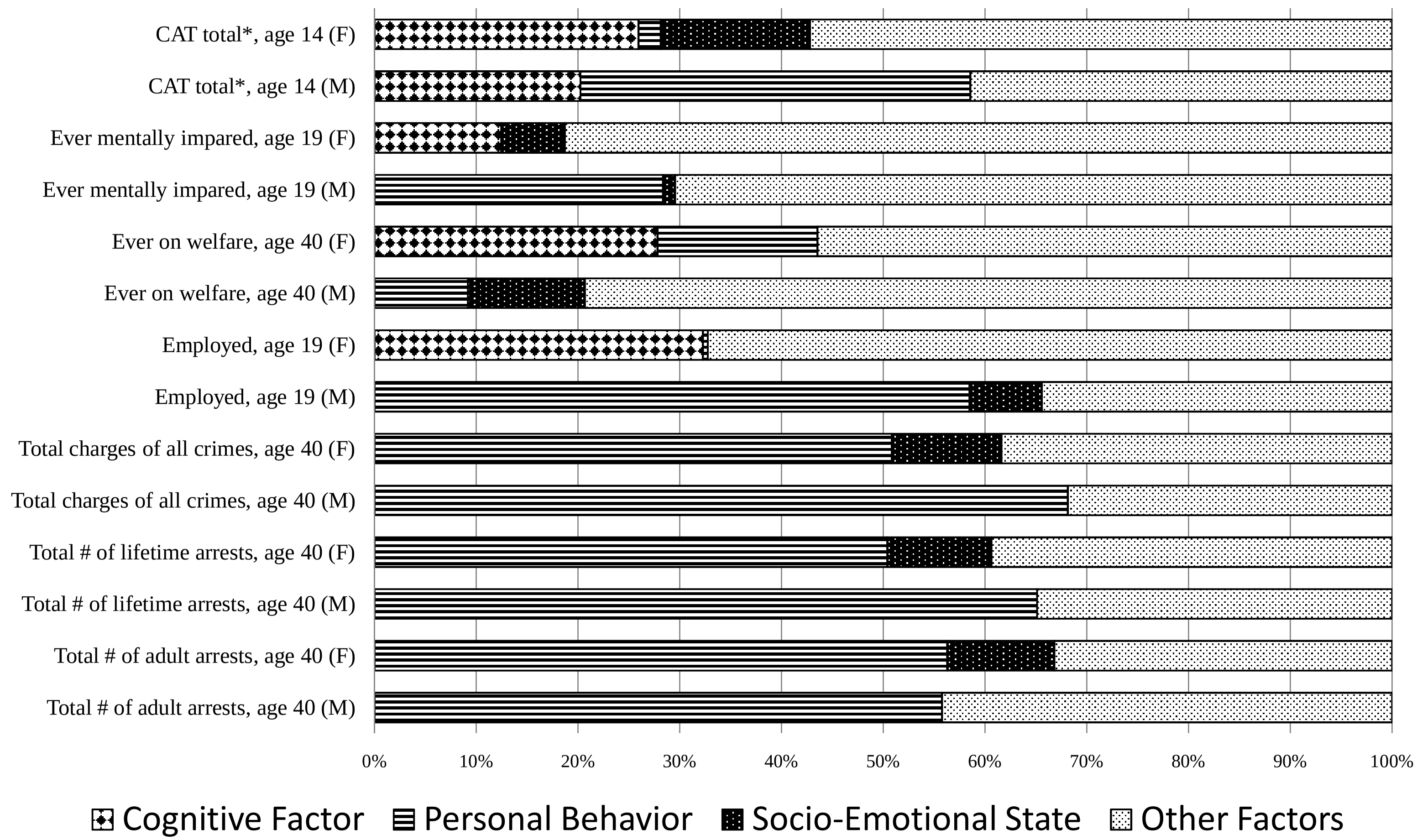

**Notes:** The total treatment effect is normalized to 100%. "(+)" and "(-)" denote positive and negative total treatment effects respectively. The figure is a slightly simplified visualization of Table 4: Small and statistically insignificant contributions of the opposite sign are set to zero. See Web Appendix I for detailed information about the simplifications made to produce the figure. Calculations are based on the Perry data. (*) "CAT total" denotes "California Achievement Test total score." Although Cognitive share estimate for CAT is about 20%, it is not statistically significant at the 10% level (one sided $p$=0.271) unlike the Personal Behavior share (one sided $p$=0.079).

**Figure 5:** Decompositions of Treatment Effects, Females



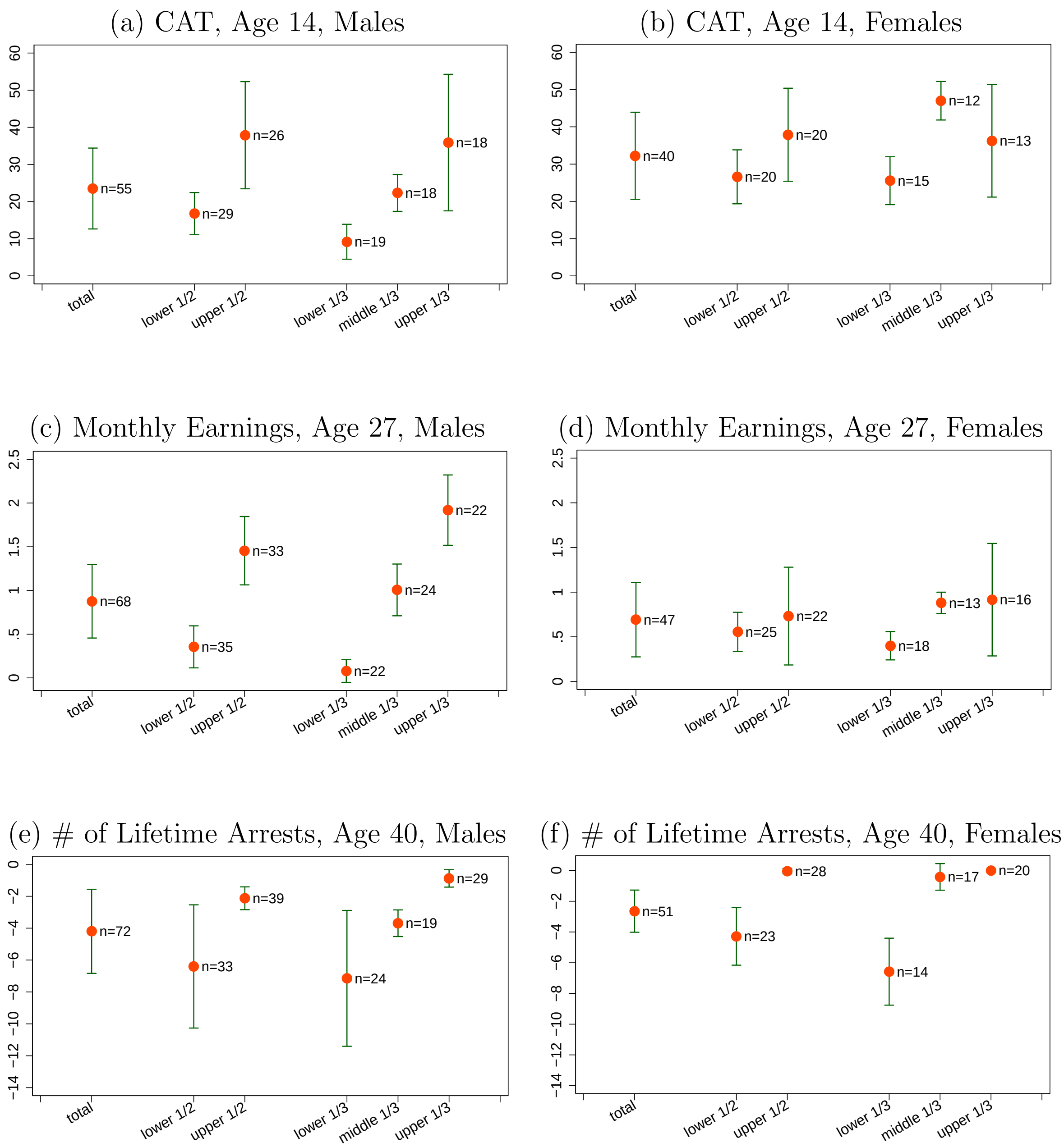

**Notes:** The total treatment effect is normalized to 100%. "(+)" and "(-)" denote positive and negative total treatment effects respectively. The figure is a slightly simplified visualization of Table 5: Small and statistically insignificant contributions of the opposite sign are set to zero. See Web Appendix I for detailed information about the simplifications made to produce the figure. Calculations are based on the Perry data. (*) "CAT total" denotes "California Achievement Test total score."

**Figure 6:** Decompositions of Treatment Effects, Gender Comparison




**Notes:** "(M)" denotes males, and "(F)" denotes females. Only treatment effects that are statistically significant for both genders are displayed. The total treatment effect is normalized to 100%. "(+)" and "(-)" denote positive and negative total treatment effects respectively. The figure is a slightly simplified visualization of Tables 4 and 5: Small and statistically insignificant contributions of the opposite sign are set to zero. See Web Appendix I for detailed information about the simplifications made to produce the figure. Calculations are based on the Perry data. (*) "CAT total" denotes "California Achievement Test total score." Although CAT's Cognitive share estimate is about 20% for males, it is not statistically significant (one sided $p$=0.271) unlike the CAT's Personal Behavior share (one sided $p$=0.079).

**Figure 7:** Quantile Treatment Effects




**Notes:** Differences in means (dots), standard errors (bars) and sample sizes ($n$) are shown. One standard error is shown in each direction. Differences in means are shown for the full sample (denoted as "total") and for various quantile groups. Earnings are in thousands of USD, US CPI-adjusted to year 2006.

**Table 1:** Descriptive Statistics

| | Variable | Age | Males | | | | Females | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Control | | Treatment | | Control | | Treatment | |
| | | | Mean | Std Err. | Mean | Std Err. | Mean | Std Err. | Mean | Std Err. |
| Background | Mother's employment | 3 | 0.28 | *(0.07)* | 0.06 | *(0.04)* | 0.35 | *(0.10)* | 0.12 | *(0.07)* |
| | Father's presence in the family | 3 | 0.56 | *(0.08)* | 0.45 | *(0.09)* | 0.42 | *(0.10)* | 0.68 | *(0.10)* |
| | SES index | 3 | 8.65 | *(0.22)* | 8.89 | *(0.30)* | 8.52 | *(0.24)* | 8.66 | *(0.29)* |
| IQ[(a)] | Stanford-Binet Intelligence Scale | 7 | 87.71 | *(1.78)* | 91.12 | *(1.72)* | 86.00 | *(1.82)* | 92.52 | *(2.78)* |
| | Stanford-Binet Intelligence Scale | 8 | 89.05 | *(1.86)* | 88.33 | *(2.26)* | 83.60 | *(1.78)* | 87.84 | *(2.82)* |
| | Stanford-Binet Intelligence Scale | 9 | 89.03 | *(2.16)* | 88.39 | *(1.80)* | 83.04 | *(2.13)* | 86.74 | *(2.48)* |
| Pupil Behavioral measures[(b)] | Absences or truancies | 7-9 | 3.73 | *(0.18)* | 4.03 | *(0.18)* | 3.70 | *(0.26)* | 3.93 | *(0.21)* |
| | Lying or cheating | 7-9 | 3.67 | *(0.15)* | 3.90 | *(0.18)* | 3.51 | *(0.19)* | 4.15 | *(0.19)* |
| | Steals | 7-9 | 4.35 | *(0.10)* | 4.54 | *(0.12)* | 4.28 | *(0.15)* | 4.63 | *(0.11)* |
| | Swears or uses obscene words | 7-9 | 3.72 | *(0.16)* | 4.26 | *(0.14)* | 4.17 | *(0.18)* | 4.53 | *(0.12)* |
| Socio-Emotional State measures[(b)] | Appears depressed | 7-9 | 3.52 | *(0.15)* | 3.62 | *(0.19)* | 3.46 | *(0.15)* | 3.74 | *(0.15)* |
| | Withdrawn and uncommunicative | 7-9 | 3.55 | *(0.18)* | 3.75 | *(0.16)* | 3.56 | *(0.19)* | 3.38 | *(0.20)* |
| | Friendly and well-received by other pupils | 7-9 | 3.39 | *(0.13)* | 3.68 | *(0.13)* | 3.22 | *(0.25)* | 3.59 | *(0.16)* |
| | Appears generally happy | 7-9 | 3.45 | *(0.13)* | 3.61 | *(0.15)* | 3.48 | *(0.22)* | 3.77 | *(0.18)* |
| Education | CAT[(c)] total | 7 | 86.53 | *(6.32)* | 90.87 | *(7.22)* | 80.73 | *(7.70)* | 105.22 | *(8.28)* |
| | CAT[(c)] total | 8 | 120.88 | *(8.19)* | 128.86 | *(8.52)* | 135.09 | *(8.93)* | 160.81 | *(10.14)* |
| | CAT[(c)] total | 14 | 93.14 | *(6.82)* | 116.67 | *(8.49)* | 96.72 | *(8.36)* | 128.95 | *(8.16)* |
| | Ever mentally impaired | ≤ 19 | 0.33 | *(0.08)* | 0.20 | *(0.07)* | 0.36 | *(0.10)* | 0.08 | *(0.06)* |
| | High school graduate | 19 | 0.51 | *(0.08)* | 0.48 | *(0.09)* | 0.23 | *(0.08)* | 0.84 | *(0.07)* |
| | Highest grade completed | 19 | 11.28 | *(0.21)* | 11.36 | *(0.23)* | 10.75 | *(0.31)* | 11.76 | *(0.22)* |
| Employment | Current employment | 19 | 0.41 | *(0.08)* | 0.55 | *(0.09)* | 0.15 | *(0.07)* | 0.44 | *(0.10)* |
| | Jobless for more than 1 year | 27 | 0.31 | *(0.07)* | 0.24 | *(0.08)* | 0.54 | *(0.10)* | 0.25 | *(0.09)* |
| | Monthly Income[(d)] | 27 | 1.43 | *(0.23)* | 2.31 | *(0.35)* | 1.13 | *(0.30)* | 1.82 | *(0.29)* |
| | Jobless for more than 2 years | 40 | 10.75 | *(1.82)* | 7.23 | *(1.71)* | 5.05 | *(1.96)* | 4.00 | *(1.49)* |
| Other outcomes | Married | ≤ 27 | 0.26 | *(0.07)* | 0.26 | *(0.08)* | 0.09 | *(0.06)* | 0.40 | *(0.10)* |
| | No tobacco use | ≤ 27 | 0.46 | *(0.08)* | 0.58 | *(0.09)* | 0.41 | *(0.11)* | 0.52 | *(0.10)* |
| | Ever on welfare | ≤ 40 | 0.82 | *(0.06)* | 0.67 | *(0.08)* | 0.92 | *(0.05)* | 0.76 | *(0.09)* |
| | Months in marriages | ≤ 40 | 59.82 | *(11.50)* | 81.61 | *(14.89)* | 47.83 | *(15.01)* | 87.48 | *(18.85)* |
| Crime Outcomes | # of adult arrests | ≤ 27 | 5.36 | *(0.93)* | 3.03 | *(0.73)* | 1.88 | *(0.77)* | 0.28 | *(0.12)* |
| | # of felony arrests | ≤ 27 | 0.49 | *(0.08)* | 0.48 | *(0.09)* | 0.15 | *(0.07)* | 0.00 | - |
| | Total charges (count), viol. crimes with victim cost | ≤ 40 | 1.00 | *(0.28)* | 0.48 | *(0.16)* | 0.04 | *(0.04)* | 0.04 | *(0.04)* |
| | Total charges of all crimes | ≤ 40 | 13.38 | *(2.10)* | 9.00 | *(1.94)* | 4.92 | *(1.30)* | 2.24 | *(0.53)* |
| | Total # of lifetime arrests | ≤ 40 | 12.41 | *(1.94)* | 8.21 | *(1.78)* | 4.85 | *(1.27)* | 2.20 | *(0.53)* |
| | Total # of adult arrests | ≤ 40 | 11.72 | *(1.83)* | 7.45 | *(1.51)* | 4.42 | *(1.22)* | 2.16 | *(0.53)* |
| | Total # of misdemeanor arrest | ≤ 40 | 8.46 | *(1.35)* | 5.33 | *(1.04)* | 4.00 | *(1.17)* | 2.12 | *(0.53)* |
| | Total charges (count), all crimes with victim cost | ≤ 40 | 3.08 | *(0.66)* | 1.48 | *(0.43)* | 0.31 | *(0.14)* | 0.04 | *(0.04)* |
| | Any charges (indic), all crimes with victim cost | ≤ 40 | 0.62 | *(0.08)* | 0.45 | *(0.09)* | 0.19 | *(0.08)* | 0.04 | *(0.04)* |
| | Sample size | | 39 | | 33 | | 26 | | 25 | |

**Notes:** Statistics are shown for variables used in the analysis. Standard errors are in parentheses. "–" indicates no variation in the data. (a) For detailed information about IQ tests, see Web Appendix B; (b) Measures of Pupil Behavioral and Socio-Emotional State are components of the Pupil Behavior Inventory (PBI). PBI items take values from 1 to 5, with higher number corresponding to a more socially-desirable behavior or trait. See Web Appendix C and Table C-3 for more information. "Age 7-9" stands for an average over non-missing observations at ages 7, 8 and 9; (c) CAT stands for California Achievement Test (see Web Appendix B.5); (d) 2006 US dollars in thousands, US CPI-adjusted.

**Table 2:** Description of Noncognitive Measures

| Traits | # of Measures | Age | Description of Measures |
|---|---|---|---|
| | | 7 | Stanford-Binet IQ; |
| **Cognition** | 3 | 8 | Stanford-Binet IQ; |
| | | 9 | Stanford-Binet IQ. |
| | | 7-9 | PBI absenses or truancies; |
| **Personal Behavior** | 4 | 7-9 | PBI lying or cheating; |
| | | 7-9 | PBI steals; |
| | | 7-9 | PBI swears or uses obscene words. |
| | | 7-9 | PBI appears depressed; |
| **Socio-Emotional State** | 4 | 7-9 | PBI withdrawn and uncommunicative; |
| | | 7-9 | PBI friendly and well-received by other pupils; |
| | | 7-9 | PBI appears generally happy. |

**Notes:** See Web Appendix B for a description of the Cognition measures. Personal Behavior and Socio-Emotional State are parts of the Pupil Behavior Inventory described in the Web Appendix C.

**Table 3:** Internal Consistency Reliability of Measures

| Capabilities | # of Measures | Cronbach's Alpha | | |
|---|---|---|---|---|
| | | Males | Females | Overall |
| **Cognitive** | 3 | 0.838 | 0.913 | 0.882 |
| **Personal Behavior** | 4 | 0.730 | 0.794 | 0.755 |
| **Socio-Emotional State** | 4 | 0.880 | 0.868 | 0.838 |

**Notes:** Cronbach's alphas (Cronbach, 1951) are calculated. Calculations are based on the Perry data.

## **Table 4:** Decompositions of Treatment Effects, Males

| Outcome | Total program effect[(a)] | Cognitive capability effect[(b)] | Personal behavior capability effect[(c)] | Socio-emotional state effect[(d)] | Total explained effect of social capabilities[(e)] | Total explained effect[(f)] | Residual effect[(g)] | *N* |
|---|---|---|---|---|---|---|---|---|
| CAT total, | 24.858 | 5.114 | 9.678 | -0.401 | 9.277 | 14.390 | 10.468 | 49 |
| age 14 | ***0.028*** | 0.271 | ***0.079*** | 0.524 | ***0.042*** | ***0.098*** | 0.142 | |
| Employed, | 0.118 | 0.000 | 0.069 | 0.008 | 0.077 | 0.077 | 0.040 | 64 |
| age 19 | 0.201 | 0.436 | 0.170 | 0.420 | 0.130 | 0.128 | 0.356 | |
| Monthly Income, | 1.074 | -0.029 | 0.285 | -0.012 | 0.273 | 0.243 | 0.831 | 60 |
| age 27 | ***0.015*** | 0.572 | ***0.098*** | 0.567 | 0.114 | 0.205 | ***0.029*** | |
| No tobacco use, | 0.210 | -0.012 | 0.119 | -0.013 | 0.106 | 0.094 | 0.116 | 62 |
| age 27 | ***0.069*** | 0.607 | ***0.054*** | 0.682 | 0.106 | 0.174 | 0.172 | |
| # of adult arrests, | -2.652 | 0.113 | -2.067 | 0.574 | -1.493 | -1.380 | -1.272 | 64 |
| age 27 | ***0.034*** | 0.571 | ***0.017*** | 0.754 | ***0.056*** | ***0.069*** | 0.204 | |
| Jobless for more than 2 years, | -3.526 | -0.010 | -1.225 | 0.063 | -1.162 | -1.172 | -2.355 | 59 |
| age 40 | 0.104 | 0.495 | 0.139 | 0.597 | 0.194 | 0.231 | 0.191 | |
| Ever on welfare, | -0.191 | 0.002 | -0.018 | -0.022 | -0.040 | -0.038 | -0.153 | 64 |
| age 40 | ***0.031*** | 0.498 | 0.354 | 0.299 | 0.231 | 0.260 | ***0.063*** | |
| total charges (count), viol. crimes with | -0.540 | 0.017 | -0.302 | 0.022 | -0.279 | -0.262 | -0.277 | 64 |
| age 40 | ***0.091*** | 0.563 | ***0.029*** | 0.606 | ***0.047*** | ***0.087*** | 0.243 | |
| Total charges of all crimes, | -4.794 | 0.182 | -4.086 | 1.021 | -3.065 | -2.883 | -1.911 | 64 |
| age 40 | ***0.068*** | 0.580 | ***0.020*** | 0.735 | ***0.043*** | ***0.061*** | 0.282 | |
| Total # of lifetime arrests, | -4.826 | 0.182 | -3.918 | 1.010 | -2.908 | -2.727 | -2.099 | 64 |
| age 40 | ***0.051*** | 0.578 | ***0.019*** | 0.745 | ***0.044*** | ***0.062*** | 0.251 | |
| Total # of adult arrests, | -4.800 | 0.164 | -3.255 | 0.872 | -2.383 | -2.219 | -2.581 | 64 |
| age 40 | ***0.031*** | 0.578 | ***0.020*** | 0.733 | ***0.050*** | ***0.070*** | 0.169 | |
| Total # of misdemeanor arrests, | -3.812 | 0.096 | -1.895 | 0.515 | -1.379 | -1.283 | -2.529 | 64 |
| age 40 | ***0.027*** | 0.586 | ***0.025*** | 0.699 | ***0.062*** | 0.100 | 0.103 | |
| Total charges of all crimes, | -1.133 | 0.048 | -1.084 | 0.269 | -0.815 | -0.767 | -0.366 | 64 |
| with victim cost, age 40 | ***0.095*** | 0.569 | ***0.021*** | 0.745 | ***0.066*** | ***0.087*** | 0.308 | |
| Any charge of a crime, | -0.166 | 0.000 | -0.072 | -0.018 | -0.090 | -0.090 | -0.076 | 64 |
| with victim cost, age 40 | 0.105 | 0.485 | 0.163 | 0.346 | 0.107 | 0.114 | 0.271 | |

**Notes:** Estimated are the following population components of the models (1–3): (a) $E(Y_{k,1} - Y_{k,0}) - \boldsymbol{\beta}_k E(\boldsymbol{X}_1 - \boldsymbol{X}_0)$; (b) $\alpha_k^C E(\theta_1^C - \theta_0^C)$; (c) $\alpha_k^B E(\theta_1^B - \theta_0^B)$; (d) $\alpha_k^S E(\theta_1^S - \theta_0^S)$; (e) $\alpha_k^B E(\theta_1^B - \theta_0^B) + \alpha_k^S E(\theta_1^S - \theta_0^S)$; (f) $\boldsymbol{\alpha}_k E(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$; (g) $\tau_{k,1} - \tau_{k,0}$, where "$C$" stands for "Cognition," "$B$" stands for "Personal Behavior," "$S$" stands for "Socio-Emotional State." One-sided bootstrap $p$-values are in parentheses. The number of bootstrap iterations is 1000. $N$ denotes the sample size. $\boldsymbol{X}$ includes three background variables at entry: mother's employment, father's presence in the household, and family's SES. Earnings are in thousands USD, US CPI-adjusted to year 2006. "CAT total" denotes the California Achievement Test total score.

**Table 5:** Decompositions of Treatment Effects, Females

| Outcome | Total program effect[(a)] | Cognitive capability effect[(b)] | Personal behavior capability effect[(c)] | Socio-emotional state effect[(d)] | Total explained effect of social capabilities[(e)] | Total explained effect[(f)] | Residual effect[(g)] | *N* |
|---|---|---|---|---|---|---|---|---|
| CAT total, | 22.184 | 12.126 | -0.313 | 1.622 | 1.309 | 13.435 | 8.749 | 39 |
| age 7 | ***0.092*** | ***0.091*** | 0.554 | 0.329 | 0.333 | ***0.069*** | 0.287 | |
| CAT total, | 27.049 | 11.403 | 3.987 | 3.307 | 7.294 | 18.697 | 8.352 | 38 |
| age 8 | ***0.073*** | 0.112 | 0.239 | 0.277 | 0.119 | ***0.048*** | 0.342 | |
| CAT total, | 29.412 | 7.630 | 0.645 | 4.315 | 4.959 | 12.590 | 16.823 | 34 |
| age 14 | ***0.017*** | 0.143 | 0.428 | 0.234 | 0.182 | ***0.069*** | ***0.082*** | |
| High school graduate, | 0.510 | -0.012 | -0.013 | 0.066 | 0.053 | 0.041 | 0.469 | 41 |
| age 19 | ***0.009*** | 0.604 | 0.628 | 0.132 | 0.213 | 0.254 | ***0.013*** | |
| Highest grade completed, | 0.854 | -0.025 | 0.034 | 0.138 | 0.172 | 0.147 | 0.707 | 39 |
| age 19 | ***0.054*** | 0.600 | 0.429 | 0.194 | 0.174 | 0.223 | ***0.084*** | |
| Employed, | 0.227 | 0.083 | 0.001 | -0.030 | -0.029 | 0.054 | 0.173 | 41 |
| age 19 | 0.101 | 0.121 | 0.508 | 0.674 | 0.664 | 0.234 | 0.188 | |
| Jobless for more than 1 year, | -0.308 | 0.053 | -0.012 | -0.092 | -0.104 | -0.051 | -0.257 | 40 |
| age 27 | ***0.077*** | 0.808 | 0.375 | 0.131 | ***0.095*** | 0.269 | 0.111 | |
| Married | 0.399 | 0.055 | 0.066 | -0.027 | 0.039 | 0.094 | 0.305 | 38 |
| age 27 | ***0.021*** | 0.216 | ***0.091*** | 0.709 | 0.244 | 0.111 | ***0.062*** | |
| Total # of adult arrests, | -2.104 | 0.286 | -0.862 | -0.061 | -0.923 | -0.637 | -1.466 | 41 |
| age 27 | ***0.029*** | 0.868 | ***0.059*** | 0.447 | 0.109 | 0.230 | ***0.040*** | |
| Total # of felony arrests | -0.155 | -0.006 | -0.056 | 0.029 | -0.027 | -0.033 | -0.121 | 41 |
| age 27 | ***0.064*** | 0.404 | 0.107 | 0.841 | 0.296 | 0.233 | ***0.036*** | |
| Ever on welfare, | -0.209 | -0.059 | -0.033 | 0.003 | -0.031 | -0.090 | -0.120 | 41 |
| age 40 | ***0.072*** | 0.128 | 0.158 | 0.478 | 0.253 | ***0.072*** | 0.202 | |
| Months in marriages, | 44.357 | 10.288 | 15.047 | -3.704 | 11.343 | 21.631 | 22.726 | 40 |
| age 40 | 0.130 | 0.205 | ***0.080*** | 0.580 | 0.170 | ***0.067*** | 0.293 | |
| Total charges of all crimes, | -1.927 | 0.360 | -1.163 | -0.246 | -1.409 | -1.049 | -0.878 | 41 |
| age 40 | 0.160 | 0.779 | ***0.075*** | 0.339 | ***0.099*** | 0.189 | 0.275 | |
| Total # of lifetime arrests, | -1.908 | 0.333 | -1.129 | -0.229 | -1.358 | -1.025 | -0.882 | 41 |
| age 40 | 0.153 | 0.765 | ***0.076*** | 0.345 | ***0.100*** | 0.183 | 0.274 | |
| Total # of adult arrests, | -1.661 | 0.323 | -1.116 | -0.209 | -1.325 | -1.002 | -0.658 | 41 |
| age 40 | 0.192 | 0.763 | ***0.073*** | 0.363 | 0.104 | 0.189 | 0.330 | |

**Notes:** Estimated are the following population components of the models (1–3): (a) $E(Y_{k,1} - Y_{k,0}) - \boldsymbol{\beta}_k E(\boldsymbol{X}_1 - \boldsymbol{X}_0)$; (b) $\alpha_k^C E(\theta_1^C - \theta_0^C)$; (c) $\alpha_k^B E(\theta_1^B - \theta_0^B)$; (d) $\alpha_k^S E(\theta_1^S - \theta_0^S)$; (e) $\alpha_k^B E(\theta_1^B - \theta_0^B) + \alpha_k^S E(\theta_1^S - \theta_0^S)$; (f) $\boldsymbol{\alpha}_k E(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$; (g) $\tau_{k,1} - \tau_{k,0}$, where "$C$" stands for "Cognition," "$B$" stands for "Personal Behavior," "$S$" stands for "Socio-Emotional State." One-sided bootstrap draws is $p$-values are in parentheses. The number of bootstrap iterations is 1000. $N$ denotes the sample size. $\boldsymbol{X}$ includes three background variables at entry: mother's employment, father's presence in the household, and family's SES. Earnings are in thousands USD, US CPI-adjusted to year 2006. "CAT total" denotes the California Achievement Test total score.

**Table 6:** Decompositions of Treatment Effects by Gender

| Outcome | Total program effect[(a)] | Cognitive capability effect[(b)] | Personal behavior capability effect[(c)] | Socio-emotional state effect[(d)] | Total explained effect of social capabilities[(e)] | Total explained effect[(f)] | Residual effect[(g)] | *N* |
|---|---|---|---|---|---|---|---|---|
| CAT total, | 29.412 | 7.630 | 0.645 | 4.315 | 4.959 | 12.590 | 16.823 | 34 |
| females, age 14 | ***0.017*** | 0.143 | 0.428 | 0.234 | 0.182 | ***0.069*** | ***0.082*** | |
| CAT total, | 24.858 | 5.114 | 9.678 | -0.401 | 9.277 | 14.390 | 10.468 | 49 |
| males, age 14 | ***0.028*** | 0.271 | ***0.079*** | 0.524 | ***0.042*** | ***0.098*** | 0.142 | |
| Ever mentally impaired, | -0.275 | -0.039 | 0.035 | -0.019 | 0.016 | -0.023 | -0.252 | 37 |
| females, age 19 | ***0.056*** | 0.270 | 0.743 | 0.392 | 0.596 | 0.406 | ***0.058*** | |
| Ever mentally impaired, | -0.181 | 0.024 | -0.058 | -0.002 | -0.060 | -0.037 | -0.144 | 59 |
| males, age 19 | ***0.080*** | 0.612 | 0.210 | 0.461 | 0.200 | 0.381 | 0.131 | |
| Employed, | 0.227 | 0.083 | 0.001 | -0.030 | -0.029 | 0.054 | 0.173 | 41 |
| females, age 19 | 0.101 | 0.121 | 0.508 | 0.674 | 0.664 | 0.234 | 0.188 | |
| Employed, | 0.118 | 0.000 | 0.069 | 0.008 | 0.077 | 0.077 | 0.040 | 64 |
| males, age 19 | 0.201 | 0.436 | 0.170 | 0.420 | 0.130 | 0.128 | 0.356 | |
| Ever on welfare, | -0.209 | -0.059 | -0.033 | 0.003 | -0.031 | -0.090 | -0.120 | 41 |
| females, age 40 | ***0.072*** | 0.128 | 0.158 | 0.478 | 0.253 | ***0.072*** | 0.202 | |
| Ever on welfare, | -0.191 | 0.002 | -0.018 | -0.022 | -0.040 | -0.038 | -0.153 | 64 |
| males, age 40 | ***0.031*** | 0.498 | 0.354 | 0.299 | 0.231 | 0.260 | ***0.063*** | |
| Total charges of all crimes, | -1.927 | 0.360 | -1.163 | -0.246 | -1.409 | -1.049 | -0.878 | 41 |
| females, age 40 | 0.160 | 0.779 | ***0.075*** | 0.339 | ***0.099*** | 0.189 | 0.275 | |
| Total charges of all crimes, | -4.794 | 0.182 | -4.086 | 1.021 | -3.065 | -2.883 | -1.911 | 64 |
| males, age 40 | ***0.068*** | 0.580 | ***0.020*** | 0.735 | ***0.043*** | ***0.061*** | 0.282 | |
| Total # of lifetime arrests, | -1.908 | 0.333 | -1.129 | -0.229 | -1.358 | -1.025 | -0.882 | 41 |
| females, age 40 | 0.153 | 0.765 | ***0.076*** | 0.345 | ***0.100*** | 0.183 | 0.274 | |
| Total # of lifetime arrests, | -4.826 | 0.182 | -3.918 | 1.010 | -2.908 | -2.727 | -2.099 | 64 |
| males, age 40 | ***0.051*** | 0.578 | ***0.019*** | 0.745 | ***0.044*** | ***0.062*** | 0.251 | |
| Total # of adult arrests, | -1.661 | 0.323 | -1.116 | -0.209 | -1.325 | -1.002 | -0.658 | 41 |
| females, age 40 | 0.192 | 0.763 | ***0.073*** | 0.363 | 0.104 | 0.189 | 0.330 | |
| Total # of adult arrests, | -4.800 | 0.164 | -3.255 | 0.872 | -2.383 | -2.219 | -2.581 | 64 |
| males, age 40 | ***0.031*** | 0.578 | ***0.020*** | 0.733 | ***0.050*** | ***0.070*** | 0.169 | |

**Notes:** Estimated are the following population components of the models (1–3): (a) $E(Y_{k,1} - Y_{k,0}) - \boldsymbol{\beta}_k E(\boldsymbol{X}_1 - \boldsymbol{X}_0)$; (b) $\alpha_k^C E(\theta_1^C - \theta_0^C)$; (c) $\alpha_k^B E(\theta_1^B - \theta_0^B)$; (d) $\alpha_k^S E(\theta_1^S - \theta_0^S)$; (e) $\alpha_k^B E(\theta_1^B - \theta_0^B) + \alpha_k^S E(\theta_1^S - \theta_0^S)$; (f) $\boldsymbol{\alpha}_k E(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$; (g) $\tau_{k,1} - \tau_{k,0}$, where "$C$" stands for "Cognition," "$B$" stands for "Personal Behavior," "$S$" stands for "Socio-Emotional State." One-sided bootstrap $p$-values are in parentheses. The number of bootstrap iterations is 1000. $N$ denotes the sample size. $\boldsymbol{X}$ includes three background variables at entry: mother's employment, father's presence in the household, and family's SES. Earnings are in thousands USD, US CPI-adjusted to year 2006. "CAT total" denotes the California Achievement Test total score.