

# Investing in our Young People\*

Flavio Cunha

University of Chicago

James J. Heckman

University of Chicago,

University College Dublin and

The American Bar Foundation

Embargoed for 12:01am on November 15, 2006

\*This paper was funded by a grant from the National Institutes of Health (R01HD43411). Additionally, we wish to thank the American Bar Foundation for their support.

## Part I

# Introduction

The study of human skill development is no longer handicapped by the taboo that once made it impermissible to discuss differences among people. It is well documented that individuals are very diverse in a variety of abilities, that these abilities account for a substantial amount of the interpersonal variation in social and economic success, and that this diversity is already apparent at an early age. The family plays a powerful role in shaping these abilities, contributing genetic endowments and prenatal and postnatal environments which interact to determine the abilities, behaviors, and talents of children. Some families do not function well, with detrimental consequences for their children. From a variety of studies, we know that it is possible to partially compensate for adverse environments if high-quality supplements are made sufficiently early in children's lives. The most effective supplements supply family resources to young children from disadvantaged environments. Successful supplements also follow up early interventions with a balanced portfolio of later interventions. Since the family is the fundamental source of inequality in American society and advantaged families spend substantial resources on their children, programs that target children from disadvantaged families have the greatest promise.

This paper develops econometric models of skill formation that distill the essence of recent empirical findings from the literature on child development. The goal is to provide a theoretical framework for interpreting the evidence from a large empirical literature, for guiding the next generation of empirical studies, and for formulating factually based policy.

Recent empirical research has substantially improved our understanding of how skills and abilities are formed over the life cycle. The early human capital literature (Becker, 1964) viewed human capital as a rival explanation for human ability in explaining earnings. It emphasized that acquired human capital could explain many features of earnings distributions and earnings dynamics that models of earnings determined by innate and invariant cognitive ability could not. More recent models (e.g. Aiyagari, Greenwood, and Seshadri, 2002; Becker and Tomes, 1979, 1986; Ben-Porath, 1967; Griliches, 1977) emphasize that innate ability is an input to the skill formation process, although its effect on human capital accumulation is ambiguous. More innate ability could lead to less schooling if all schooling does is to teach what an

able person could learn without formal instruction. On the other hand, more innate ability might make learning easier and promote schooling. The signaling literature (Spence, 1973; Stiglitz, 1975) focused on the latter interpretation in developing models of education where higher levels of schooling signal higher innate ability. In its extreme form, this literature suggested that there was no learning content in schooling.

The entire literature assumes that ability is an innate, scalar, age-invariant measure of cognitive skill. This early point of view still prevails in most quarters of economics. Except for work by Marxist economists (see, e.g. Bowles and Gintis, 1976; Edwards, 1976), noncognitive traits like motivation, persistence, time preference, and self control were neglected in empirical research and treated as “soft skills,” peripheral to the study of educational and labor market outcomes.

The recent economic literature on family influence on child outcomes focuses on family income constraints and heritability as the principal sources of parental influence on child development. Becker and Tomes (1979, 1986) initiated a large literature that emphasized the importance of credit constraints and family income on the schooling and earnings of children. Important developments of this work by Benabou (2000, 2002), Aiyagari, Greenwood, and Seshadri (2002), Caucutt and Kumar (2003), Hanushek, Leung, and Yilmaz (2004), and Seshadri and Yuki (2004), emphasize the role of credit constraints and parental altruism in forming the skills of children. In this line of research, ability is treated as determined by genetic factors. The life cycle of the child at home is collapsed into a single period so that there is no distinction between early and late investments in children. Becker and Tomes (1986) show that there is no trade-off between equity and efficiency in making government transfers directed toward credit-constrained families because the return to human capital investment in children from such families is high due to the presence of credit constraints. We show that their insight holds true for early period investments in a multi-period model of child investment, but not for investments in later periods. It is important for studying the economics of skill formation to disaggregate the life cycle of the child and distinguish infancy, early schooling, and adolescent outcomes.

Recent research, summarized in Heckman (2000), Carneiro and Heckman (2003), and Cunha, Heckman, Lochner, and Masterov (2006) presents a richer picture of schooling, life cycle skill formation and earnings determination. It recognizes the importance of both cognitive and noncognitive abilities in explaining schooling and socioeconomic success. These abilities are produced by the family and by personal actions. The role of the mother is especially important. Both genes and environments are involved in producing

these abilities. Environments affect genetic expression mechanisms (see, e.g. Turkheimer, Haley, Waldron, D’Onofrio, and Gottesman, 2003). This interaction has important theoretical and empirical implications for policies to promote skill. It suggests an important role for environment-enriching policies in fostering the production of human skills.

In the light of a substantial body of recent research, the traditional sharp distinction between acquired skills and genetically determined cognitive ability maintained in the human capital literature is no longer tenable. Abilities are multiple in nature and are both cognitive and noncognitive. Measured cognitive ability is susceptible to environmental influences, including *in utero* experiences. So is measured noncognitive ability. There are genetic components to both. We have come to understand that achievement tests used to monitor performance in school and to determine acceptance into the military are not the same as IQ tests. Achievement test scores are affected by IQ, schooling inputs, and noncognitive skills. Noncognitive abilities such as motivation, self-discipline, and time preference—associated with the development of the prefrontal cortex—are also affected by environmental influences. They are more malleable at later ages than IQ. Achievement test outcomes can be influenced until very late ages and are affected by both cognitive and noncognitive skills. Noncognitive abilities and cognitive abilities affect schooling attainment and performance, and a wide array of behaviors (Cunha and Heckman, 2006; Heckman, Stixrud, and Urzua, 2006). Abilities have an acquired character although they differ in their malleability at different ages.

We characterize the human skill formation process in the following fashion. Skills and abilities are used interchangeably throughout this paper because both are affected by environments, investment and genes. Agents possess a vector of abilities at each age. These abilities—or skills—are multiple in nature and range from pure cognitive abilities (e.g. IQ) to noncognitive abilities (patience, self control, motivation, temperament, time preference). Achievement test scores are affected by cognitive, noncognitive and environmental inputs. These abilities are used with different weights in different tasks in the labor market and in social life more generally.

The human skill (ability) formation process is governed by a multistage technology. Each stage corresponds to a period in the life cycle of a child. Inputs or investments at each stage produce outputs at that stage. Unlike the Ben Porath model (1967), in our models qualitatively different inputs can be used at different stages and the technologies may be different at different stages.<sup>1</sup> The outputs at each stage

---

<sup>1</sup>Heckman, Lochner, and Taber (1998) generalize and estimate the Ben Porath model by allowing the technology producing

are the levels of each skill achieved at that stage. Some stages of the technology may be more productive in producing some skills than other stages, and some inputs may be more productive at some stages than at other stages. Those stages that are more productive in producing certain skills are called “sensitive periods” for those skills. If one stage alone is effective in producing a skill (or ability) it is called a “critical period” for that skill.

An important feature of the technology of skill formation is that the skills produced at one stage augment the skills attained at later stages. This is *self-productivity*. Skills acquired in one period persist into future periods. In addition, skills are self-reinforcing. A second important feature of the skill formation process is *complementarity*. Skills produced at one stage raise the productivity of investment at subsequent stages. For example, self-control and emotional security may reinforce intellectual curiosity and promote more vigorous learning of cognitive skills (see Duncan, Claessens, and Engel, 2004). In a multistage technology, complementarity implies that levels of skill investments at different ages bolster each other. They are synergistic. Complementarity also implies that early investment has to be followed up by later investment in order for the early investment to be productive. Together, complementarity and self-productivity produce multiplier effects which explain how skills beget skills and abilities beget abilities.

Complementarity, self-productivity of skills and multiplier effects imply an equity-efficiency trade-off for late child investments but not for early investments. These features of the technology of skill formation have consequences for the design and evaluation of public policies toward families.

The plan of this paper is as follows. Part II presents the evidence. Part III presents simple formal models that summarize the evidence using economic theory. In Part IV we report on the results from recent research that estimates the new economic models of skill formation. In Part V we test the model by conducting out-of-sample prediction checks and we use the best-performing models to simulate the impact of different policies aimed at reducing poverty. We conclude Part V with a discussion demonstrating that the future of the U.S. economy is linked to the quality of American youth. If we fail to produce a skilled, educated workforce, our economic performance in the future will not be as strong as in the past. Part VI concludes.

---

schooling human capital to be different from the technology producing post-school investment. Su (2004) and Cardak and Givon (2004) develop multistage models of secondary and postsecondary schooling choices focusing on determinants of progression through school. However, their emphasis is on later stages of the life cycle, not the early years.

## Part II

# A Summary of the Empirical Evidence on Life Cycle Skill Formation

## 1 Human Capital Accumulation

Skill formation is a dynamic process. The skills and abilities acquired in one stage of the life cycle affect the productivity of learning in the next stage. We can think of human capital as a combination of different types of skills and abilities. It is now well established that cognitive ability is an important determinant of schooling and labor market outcomes.<sup>2</sup> At the same time, noncognitive abilities, although harder to measure, play an important role as well.<sup>3</sup> As emphasized in recent studies of child development (e.g. Shonkoff and Phillips, 2000), different abilities are formed and shaped at different stages of the life cycle. Empirical evidence from human and animal species tells us that when the opportunities for formation of these abilities are missed, remediation can be costly, and full remediation prohibitively costly (Cameron, 2004; Knudsen, 2004; Knudsen, Heckman, Cameron, and Shonkoff, 2006). These findings highlight the need for economists to take a comprehensive view of skill formation over the life cycle.

Cognitive ability is only one aspect of human skill. It is necessary for success in life, but for many aspects of performance in social life it is not enough. Noncognitive abilities also matter for success both in the labor market and in schooling. Even when early childhood interventions do not boost IQ, they appear to improve noncognitive skills (motivation, persistence, and the like), with substantial effects on schooling, labor market outcomes, and behavioral outcomes such as teenage pregnancy and participation in criminal activities. They raise achievement test scores, which can be influenced by schooling (and other inputs), even when they do not boost IQ. In light of this evidence, the neglect of noncognitive ability in evaluating human capital interventions and in analyzing the skill formation process is not justified. We summarize the evidence on the importance of noncognitive skills in Section 2. Both cognitive and noncognitive skills or

---

<sup>2</sup>See Heckman (1995).

<sup>3</sup>See the evidence in Heckman, Stixrud, and Urzua (2006), and the papers they cite.

abilities are affected by families and schools. They differ in their malleability over the life cycle. Differences in levels of cognitive and noncognitive skills by family income and family background emerge early and persist. If anything, schooling in the early grades widens these differences. However, most of the gaps in these skills that are found in adulthood emerge before schooling begins.

## 2 The Evidence on the Importance of Noncognitive Skills

Much of the neglect of noncognitive skills in analyses of earnings, schooling, and other life outcomes is due to the lack of reliable measurements of them. Many different personality traits are lumped into the category of noncognitive skills. Psychologists have developed batteries of tests to measure these skills (Sternberg, 1985). Companies use personality tests to screen workers, but they are not yet widely used to ascertain college readiness or to evaluate the effectiveness of schools or reforms of schools. The literature on cognitive tests shows that one dominant factor (“*g*”) summarizes cognitive tests and their effects on outcomes. No single factor has emerged as dominant in the literature on noncognitive skills and it is unlikely that one will ever be found, given the diversity of traits subsumed under the category of noncognitive skills. Heckman, Stixrud, and Urzua (2006), test and reject the “*g*” theory of noncognitive skills.

Studies by Bowles and Gintis (1976), Edwards (1976), and Klein, Spady, and Weiss (1991) demonstrate that job stability and dependability are the traits most valued by employers as ascertained by supervisor ratings and questions of employers, although they present no direct evidence of the effects of these traits on wages and educational attainment. Perseverance, dependability and consistency are the most important predictors of grades in school (Bowles and Gintis, 1976).

Self-reported measures of persistence, self-esteem, optimism, future orientedness, and the like are now collected in major data sets, and some recent papers discuss estimates of the effects of these measures on earnings and schooling outcomes (see Bowles and Gintis, 1976; Duncan, Claessens, and Engel, 2004). These studies shed new light on the importance of noncognitive skills for success in social life. Yet these studies are not without controversy. For example, *ex post* assessments of self-esteem may be as much the consequence as the cause of the measures being investigated.

Heckman and Rubinstein (2001) avoid the problems inherent in these *ex post* assessments by using evidence from the GED testing program in the United States to demonstrate the quantitative importance

of noncognitive skills in determining earnings and educational attainment. The GED program is a second-chance program that administers a battery of cognitive tests to self-selected high school dropouts to determine whether or not their level of academic attainment is equivalent to that of high school graduates.

The GED examination is successful in psychometrically equating GED test takers with ordinary high school graduates who do not go on to college. Recipients are as smart as ordinary high school graduates who do not go on to college, where cognitive ability is measured by an average of cognitive components of the AFQT or by the first principal component (“ $g$ ”) derived from the components. According to these same measures, GED recipients are smarter than other high school dropouts who do not obtain a GED (see Heckman and Rubinstein, 2001). In the raw data, GED recipients earn more than ordinary high school dropouts, have higher hourly wages, and finish more years of high school before they drop out. This is entirely consistent with the literature that emphasizes the importance of cognitive skills in determining labor market outcomes.

When measured ability is controlled for, however, GED recipients earn the same as or less than other dropouts. Heckman and Rubinstein (2001) note that noncognitive skills play an important role in this gap. GEDs have higher cognitive skills than dropouts but exhibit the same problems of self control and self discipline exhibited by dropouts, and on some behaviors are worse than other dropouts.

Heckman, Stixrud, and Urzua (2006) present evidence that both cognitive and noncognitive skills affect schooling and the returns to schooling.<sup>4</sup> They analyze the changes in the probabilities of various outcomes that arise from changing cognitive or noncognitive abilities. Figures 1A and 1B, taken from their study, show that higher levels of both cognitive and noncognitive skills are associated with lower rates of dropping out of high school. For many outcome measures, increasing noncognitive ability over the same decile range as cognitive ability has a greater effect on outcomes than increasing cognitive ability over the same decile

---

<sup>4</sup>Cognitive and noncognitive abilities are estimated using a two-factor model and the NLSY79 data. The cognitive skill is identified by using a subset of five Armed Forces Vocational Aptitude Battery (ASVAB) tests (word knowledge, paragraph comprehension, numerical operations, coding speed and mathematics knowledge). The noncognitive factor is identified using the Rosenberg Self-Esteem and Rotter Locus of Control scales. The Rosenberg scale contains ten statements of self-approval and disapproval with which respondents are asked to strongly agree, agree, disagree or strongly disagree. A high score indicates a high self-approval rating. The Rotter scale is based on four questions about the extent to which respondents believe themselves to have control over the various domains of their lives. A higher score indicates more control over one’s life. All tests were administered in 1979-81, when the respondents were 14-24 years old. The estimation of the model is carried out using an MCMC routine. Heckman, Stixrud and Urzua use only the young sample to analyze the data (the scores are measured at least 3-4 years before the outcomes). They also show results from other data sets where the separation between the age of the test and the outcome is more substantial, and they find very similar results. They apply the method developed in Hansen, Heckman, and Mullen (2004) to account for spurious feedback between outcomes and test scores.

range. These effects are not always uniform across genders.<sup>5</sup>

Increasing noncognitive ability to the highest level reduces the probability of being a high school dropout to virtually zero for females with average cognitive ability (see Figure 1B).<sup>6</sup> This effect is especially pronounced at the bottom of the distribution (going up from the bottom fifth). The effect is less strong for males. Both cognitive and noncognitive skills are strong predictors of who graduates from a four year college but the effects of noncognitive skills are stronger for females (see Figures 1C and 1D). Increases in both types of ability have the same effect on reducing the likelihood of spending time in jail by age 30 for males (see Figure 1E).<sup>7</sup> Figures 1F and 1G show strong effects of both cognitive and noncognitive skills on smoking. Here there is a larger effect for males of increasing noncognitive ability. Figure 1H shows the strong effect of both cognitive and noncognitive skills on non-marital pregnancy. For this outcome both cognitive and noncognitive ability are important.<sup>8</sup> Higher levels of noncognitive skills promote success on achievement tests even when they do not affect IQ. This effect operates because noncognitive skills affect schooling and schooling raises measured achievement. (Hansen, Heckman, and Mullen, 2004).

Current systems of evaluating educational reforms are based predominantly on changes in scores on cognitive tests. These tests capture only one of the many skills required for a successful life (see Heckman, 1999). A more comprehensive evaluation of educational systems would account for their effects on producing the noncognitive traits that are also valued in the market. There is substantial evidence that mentoring and motivational programs oriented toward disadvantaged teenagers are effective. We review this evidence in Section 5.

Much of the effectiveness of early childhood interventions comes from boosting noncognitive skills and from fostering motivation.<sup>9</sup> More motivated children are more likely to stay in school and have higher achievement tests. Our analysis suggests that social policy should be more active in attempting to alter noncognitive traits, including values, especially for children from disadvantaged environments who receive poor discipline and little encouragement at home. This more active social policy approach would include mentoring programs and stricter enforcement of discipline in the schools. Although such programs are

---

<sup>5</sup>Heckman, Stixrud, and Urzua (2006) show how this nonuniformity in the effects of cognitive and noncognitive skills on outcomes across genders can explain the differential effectiveness of early intervention programs across genders.

<sup>6</sup>Heckman, Stixrud, and Urzua (2006) show the same patterns apply to college attendance.

<sup>7</sup>Incarceration is not an important phenomenon for females.

<sup>8</sup>Heckman, Stixrud, and Urzua (2006) show the same pattern for other reproductive outcomes, such as marital childbearing.

<sup>9</sup>See Karoly, Greenwood, Everingham, Hoube, Kilburn, Rydell, Sanders, and Chiesa (1998), Currie and Blau (2006), and Heckman (2000) for comprehensive reviews of the literature.

controversial, they are likely to be effective and to produce substantial saving to society from reduced pathological behavior (see our discussion in Section 5.1).

The evidence shows that both cognitive and noncognitive skills as measured at adolescent years can predict success in a large set of adult outcomes. In what follows we show that the gaps in cognitive and noncognitive skills across family income groups are already present early in the life of the child.

### 3 Early Test Score Differentials

Important differences in the ability of children across family types appear at early ages and persist. Figure 2A plots average percentile ranks<sup>10</sup> on the Peabody Individual Achievement Test in Math (PIAT Math) by age for different quartiles of family income. This test is a measure of age-appropriate math knowledge. There are large gaps by the time children enter school. The gaps in ranks across income quartiles remain relatively stable as children develop. Such gaps also appear in other test scores, although for some test scores they widen slightly.<sup>11</sup> Just as income gradients in schooling participation rates are evident, racial differences in early test scores also emerge. Figure 2B presents evidence on the emergence of racial gaps in ranks on the PIAT Math Test.

Ability affects schooling participation and affects wages as we document above. It is shaped early in life. Having access to more and higher-quality resources that contribute to improving cognitive ability early in life affects skill acquisition later in life. IQ is not the same as what is measured by achievement tests. Achievement tests are affected by schooling and other environmental influences into adolescence even if IQ is not (see Hansen, Heckman, and Mullen, 2004; Heckman, Stixrud, and Urzua, 2006).

Figures 3A and 3B present the gaps in PIAT Math from the previous two figures after controlling for some main features of the child's family background. The gaps across racial and income groups are significantly reduced when we control for maternal education and cognitive ability,<sup>12</sup> and for family structure. Measured long-term family factors play a powerful role in a correlational sense. The gaps at age

---

<sup>10</sup>In constructing the graph in Figure 2A, we computed each individual's position in the distribution of test scores at each age. Then we divided individuals into different quartiles of permanent family income and computed the average percentile rank at each age. Because the scale of test scores is arbitrary, an analysis of test scores can only determine how the factors being studied shift people in the overall distribution of ability.

<sup>11</sup>For evidence on other tests, see Carneiro, Heckman, and Masterov (2005)

<sup>12</sup>Cognitive ability is measured using the Armed Forces Qualifications Test, corrected for the effect of schooling using the methodology of Hansen, Heckman, and Mullen (2004).

12 do not disappear entirely, however, when we compare the highest and lowest income quartiles or whites with blacks. The evidence from early intervention programs with randomized assignment that we discuss in Section 5 indicates that these correlational results have a causal basis. When disadvantaged children are given enriched early environments, the gaps in academic achievement test scores between advantaged and disadvantaged children can be partially remedied.

The emergence of early test score gradients is not limited to cognitive measures. At early ages, differences in children’s behavior across income and racial groups are also evident, as Figures 4A and 4B illustrate. These figures present differences in ranks on an index of Anti-Social Behavior across different income and racial groups. The Anti-Social Behavior index is based on exhibiting age-specific behaviors like cheating and telling lies, bullying and cruelty to others, not feeling sorry for misbehaving, breaking things deliberately, disobedience at school, and trouble getting along with teachers. High values of the index correspond to a higher prevalence of behavioral problems. As we discuss further in Section 2, understanding the gaps in these behavioral skills across different income and racial groups and how to eliminate them is important for understanding the determinants of economic success. Figures 5A and 5B present Anti-Social Behavior index adjusted for mother’s ability, mother’s AFQT, and broken home.<sup>13</sup> Adjusting for early family background factors substantially reduces gaps in ranks in noncognitive skills across income and racial groups. Comparing adjusted cognitive and noncognitive test scores reveals the importance of long-term factors in reducing the gaps in behavioral scores across these groups. Although noncognitive ability gaps across income and racial groups cannot be fully eliminated by a regression adjustment, controlling for mother’s ability and education, family income, and family structure significantly reduces the gaps in noncognitive abilities across these groups at both early and later ages. The experimental evidence discussed in Section 5 confirms that these findings on noncognitive skills have a causal basis. Indeed, the evidence across a variety of studies suggests that early childhood interventions affect motivation and other noncognitive skills.

This evidence suggests that strong families (those with enriched parental environments) promote cognitive, social, and behavioral skills. Weak families do not. This conclusion is consistent with a large body of

---

<sup>13</sup>We first regress the Anti-Social score on mother’s education, mother’s AFQT, and broken home at the same age at which the score is measured. We then rank individuals on the residuals of this regression and construct percentiles. We then include family income in the regression as well as the other variables mentioned above before taking the residuals and constructing the ranks.

evidence in sociology and economics (see, e.g. Duncan and Brooks-Gunn, 1997). The relevant policy issue is to determine what interventions in dysfunctional families, if any, are successful. The evidence presented in Section 5 addresses this question.

## 4 Critical Periods, Sensitive Periods, and Socioemotional Bases of Skill Formation and Remediation

Early experience exerts a profound influence on socioemotional outcomes directly, but it also interacts with genetic endowments, with consequences that are at least as important for development.<sup>14</sup> Experimental studies using animals have produced several suggestive findings that enhance our understanding of the evidence on human behavior. Knudsen, Heckman, Cameron, and Shonkoff (2006) summarize a large body of evidence from human and animal studies.

Suomi (1999) summarizes his research on the malleability of temperament. He and his colleagues selectively bred rhesus monkeys to be highly fearful. They then reassigned some of these infants to nurturing mothers, while pairing some infants of normal mothers with fearful adoptive mothers. Their results suggest that normal infants take on their foster mother's fearful characteristics. Infants born to fearful mothers assigned to nurturing mothers become even more socially precocious than their normal counterparts. They engage in autonomous exploration of their environment earlier and more frequently, and they do not display disproportionate responses to minor alarming stimuli. When they are moved into larger social groups, they are able to recruit allies and attain higher positions in the monkey hierarchy. Regardless of their genetic background, young females acquired the nurturing style of their adoptive mother with their own offspring rather than the style predicted by their genetic profile or own biological mother's behavior. These results suggest that positive early experiences can dramatically modify apparent genetic tendencies, as expressed in behavior.

Knudsen (2004) shows that early experience can modify the biochemistry and architecture of neural circuits. When such experiences operate within a *limited* time frame in the life cycle, that period is termed "sensitive." During a sensitive period, certain patterns of connectivity among neurons become stable as

---

<sup>14</sup>A twins study by Turkheimer, Haley, Waldron, D'Onofrio, and Gottesman (2003) found that in poor families, 60% of the variance in IQ is accounted for by the shared environments, and the contribution of genes is close to zero, whereas in wealthy families a nearly opposite result is found.

a result of environmental influence. This stability is environmentally adaptive.<sup>15</sup> These pathways can be altered after the sensitive period, but their plasticity is limited by the structure created during the sensitive period, i.e., it is less efficient to invest in later periods. When experience in a given period is crucial for normal development, that period is called “critical.” We formally define sensitive and critical periods in Part III. Intuitively, if late investment is a good substitute for early investment, the early years are not critical. If it is not a good substitute, then the early period is critical.

Critical periods have been extensively documented in the development of binocular vision in the cortex of mammals, auditory space processing in the midbrain of barn owls, filial imprinting in the forebrain of ducks and chickens, and song learning in the forebrain of songbirds (see Knudsen, 2004). For our purposes, the most relevant example is language acquisition and the fact that children tend to perform better in acquiring language skills than do adults, despite being more limited in most cognitive domains. Age of exposure to a language is negatively related to ultimate proficiency achieved in that language (see Newport, 2002, for a summary of the evidence). The decline in proficiency begins as early as 4 and 6, and continues until a plateau is reached in adulthood. This pattern is evident for many aspects of language proficiency, such as control over sounds as well as grammatical structure, and has been shown for both first and second languages.<sup>16</sup> However, not all aspects of language acquisition are equally sensitive. Newport (2002) cites evidence that the acquisition of vocabulary and semantic processing can be accomplished relatively easily even in adulthood, while the more formal dimensions of language, such as syntax, phonology, and morphology, are less easily acquired. These differences are apparent even on a neurological level. In short, both critical and sensitive periods are features of language learning.

Other types of social behavior are characterized by sensitive and critical periods. Independent research by Cameron (2004) suggests that development of normal social behavior in infant rhesus monkeys can be disrupted by removing the mother from the social group. When mothers and infants are separated when the infants are one week old, their subsequent adult behavior is profoundly antisocial, anxious, and aggressive. When the disruption takes place at a later age, the effects are qualitatively different and their

---

<sup>15</sup>Knudsen (2004) argues that experience provides information about the individual and his environment that cannot be predicted accurately and, therefore, cannot be encoded genetically. This may explain why the early experience of deprivation may result in maladaptive development and corresponding behavior. In some sense, the adaptation may only be adaptive locally, rather than globally.

<sup>16</sup>The age-of-exposure effect appears even in the grammatical skills of deaf adults who learn sign language. See Pinker (1994) and Newport (2002) for more on this topic.

severity declines with age at separation. The impact on the youngest monkeys can be offset by pairing them with an experienced mother, but the degree of catch-up decreases with the age at which the “foster” placement takes place. Remediation is possible, though its timing is crucial.

The monkeys who are emotionally secure explore more and learn more. This evidence shows how noncognitive skills feed into the formation of cognitive skills. It helps to explain how the Perry Preschool Program, discussed in Section 5, which did not raise IQ, but raised noncognitive skills, affected achievement test outcomes. We formalize the notion of critical and sensitive periods in Part III and in the Appendix to this paper. Closely related is the concept of a “bottleneck” period. If skills at one stage of the life cycle are not formed at a sufficiently high level, it is difficult to proceed to excellence at the next stage. The “Leontief” technology discussed in Section 6 formalizes this point.

We turn next to an analysis of the evidence on the effectiveness of specific policies in supplementing the environments of disadvantaged children.

## **5 What is Known About Specific Policies to Foster Ability and Skill?**

### **5.1 Early Interventions**

Karoly, Greenwood, Everingham, Hoube, Kilburn, Rydell, Sanders, and Chiesa (1998), Currie (2001) and Currie and Blau (2006) present comprehensive surveys of numerous preschool intervention programs targeted toward disadvantaged populations and their measured effects. The programs they analyze vary, both in terms of age of enrollment and age of exit. The effects are generally consistent, although in some cases they are quite small.<sup>17</sup> Generally, performance of children in school is improved in terms of less grade repetition, more graduation and higher test scores. Unfortunately, many of the evaluations of these

---

<sup>17</sup>For example, Currie and Thomas (2000) show that test score gains of participants in the Head Start program tend to fade completely for blacks but not for whites. Their paper suggests that one reason may be that blacks attend worse schools than whites, and therefore blacks are not able to maintain initial test score gains. However, Heckman, Larenas, and Urzua (2004) dispute this finding. They show that schooling quality differences, which are substantial across ethnic groups, have only a slight effect on the levels or rates of growth in test scores. Garces, Thomas, and Currie (2002) find comparable results. The *Mathematica* evaluation of Early Head Start by Love, Eliason-Kisker, Ross, Schochet, Brooks-Gunn, Paulsell, Boller, Constantine, Vogel, Sidle Fuligni, and Brady-Smith (2002) shows very modest effects as well. However, Head Start is a considerably less intensive program, which may explain why it has limited consequences for the developmental trajectories of disadvantaged children.

programs do not follow children into late adolescence or adulthood. Interventions at younger ages seem to produce larger effects.<sup>18</sup>

Three programs have long-term follow-ups, and we focus on them here. They all targeted high-risk children from disadvantaged families. The first is the High/Scope Perry Preschool, a half-day program on a small scale in the Ypsilanti, MI public schools. Children were typically enrolled at age 4 and stayed in the program for two years. It was an experiment with a sample size of 123 and follow-up to age 40. The Abecedarian program, the second one we consider, was a full-day, year-round educational child care program in Chapel Hill, NC. Children entered around the age of 4 months and continued until age 5 years. Half of all children were then enrolled in a school-age program until age 8. It was evaluated by randomization and has 111 participants, and students are followed to age 21.

The final program we consider is the Chicago Child-Parent Centers (CPC), a half-day program during the school year and full-time for six weeks during the summer, conducted on a large scale in the Chicago public schools. It was evaluated by a non-experimental method (matching) and has a sample of about 1,500 children. All three programs had some sort of parental involvement component, and data collection is ongoing.

The programs differ by duration and child age at entry. Abecedarian started with young children in the first months of life. Perry and the CPC program start with older children, 3-5 years old. The programs differ in intensity. For some programs the comparison group received some supplementary resources relative to ordinary children, and for others they did not. Some comparison group members for some programs attend alternative preschool and kindergarten programs.

### **5.1.1 Perry Preschool Experiment**

The Perry preschool experiment was an intensive preschool program that was administered to 65 randomly selected black children who were enrolled in the program over 5 different waves between 1962 and 1967. All the children came from Ypsilanti, MI. A control group of roughly the same size provides researchers with an appropriate benchmark to evaluate the effects of the preschool program.

The experimental group assignment was performed in the following way. Candidate families were

---

<sup>18</sup>Morris, Duncan, and Clark-Kauffman (2005) find that the biggest impact of a parental wage-subsidy intervention on children's achievement is for preschool children.

identified from a census of the families of the students attending the Perry school at the date of operation of the program, neighborhood group referrals and door to door canvassing. Poor children who scored between 75 and 85 on the standard Stanford-Binet IQ test were randomly divided into two undesignated groups.<sup>19</sup> The children were then transferred across groups to equalize the socioeconomic status, cognitive ability (as measured by the IQ test) and gender composition of the samples. Finally, a coin was tossed to determine which group received the treatment and which did not. Initially the treatment and control groups included 64 children each, but the actual treatment and control groups contained 58 and 65 children, respectively.<sup>20</sup>

Children entered the Perry School in five waves, starting with wave zero (of four-year-olds) and wave one (of three-year-olds) in 1962, then waves two, three and four (of three-year-olds) entered in each subsequent year through 1965. The average age at entry was 42.3 months. With the exception of wave zero, treatment children spent two years attending the program. In the final year of the program, 11 three-year-olds who were not included in the data attended the program with the 12 4-year-olds who were. About half of the children were living with two parents. The average mother was 29 years old and completed 9.4 years of school.

The treatment consisted of a daily  $2\frac{1}{2}$  hour classroom session on weekday mornings and a weekly ninety minute home visit by the teacher on weekday afternoons to involve the mother in the child's educational process. The length of each preschool year was 30 weeks, beginning in mid-October and ending in May. Ten female teachers filled the four teaching positions over the course of the study, resulting in an average child-

---

<sup>19</sup>Poverty status was determined by a formula that considered rooms per person in the child's household, parental schooling and occupational level. The IQ range was labeled as "borderline educable mentally retarded" by the state of Michigan at the time of the experiment. Only children without an organic mental handicap were included in the study.

<sup>20</sup>Some aspect of the assignment was clearly nonrandom and this has led some to call the Perry results into question. First, younger children were assigned to the same group as their older siblings. Two treatment children were transferred to the control group because their mothers were not able to participate in any classes or home visits because they were employed far from home. Four treatment children left the program before completing the second year of preschool when their families relocated, and one control child died. Thus, the final sample consisted of 123 children. The 123 children in the sample came from 100 families. In the control group, 41 families contributed 1 child each, and 12 families contributed 2 children each. In the treatment group, 39 families contributed 1 child apiece, 6 families contributed 2 children apiece, 1 family contributed 3 and another 4 children. Assigning younger siblings to the same group effectively made the family, rather than the individual, the unit of analysis. Still, it is difficult to argue that assigning siblings at random would have been a better strategy. So-called spillovers to the control siblings from home visits would have been one possible source of bias since mothers cannot be expected to treat siblings in accordance with their experimental status. Another potential source of bias is spillover from one sibling to another. In any case, differences in background characteristics between the two experimental groups are virtually nonexistent, with the exception of much higher rates of maternal employment at program entry in the treatment group.

teacher ratio of 5.7 for the duration of the program.<sup>21</sup> All teachers were certified to teach in elementary, early childhood or special education.<sup>22</sup> If it were administered today, the Perry preschool program would cost approximately \$9,785 per participant per year in 2004 dollars. That compares to about \$7,500 per pupil per year for ordinary public education.

### 5.1.2 Abecedarian Project

The Abecedarian Project recruited 111 children born between 1972 and 1977 whose 109 families scored high on the High Risk Index.<sup>23</sup> It enrolled families and intervened in the lives of children beginning a few months after birth. Enrollment was based on the characteristics of the families more than on the characteristics of the children, as in the Perry program. Virtually all of the children were Black, and their parents had low levels of education, income, cognitive ability and high levels of pathological behavior. The children were screened for mental retardation. 76% of the children lived in a single parent or multigenerational household. The average mother in this group was less than 20 years old, completed 10 years of schooling and had an IQ of 85. There were 4 cohorts of about 28 students each. By the time they were 6 weeks old, the children were assigned randomly to either a preschool intervention or a control group. The mean age of entry was 4.4 months. At age 5, just as they were about to enter kindergarten, all of the children were reassigned to either a school age intervention through age 8 or to a control group. This yielded 4 groups: children who experienced no intervention at all, those who experienced an intervention when they were young, those who experienced it when they were older, and finally those who enjoyed a high-quality intervention throughout their whole childhood. The children were followed up until age 21.

The Abecedarian intervention was more intensive than the Perry one. The preschool program was a year-round, full-day intervention. The initial infant-to-teacher ratio was 3:1, though it grew to a child-to-teacher ratio of 6:1 as the kids progressed through the program. Infants in the control group received an iron-fortified formula for 15 months and diapers as needed to create an incentive for participation. Many

---

<sup>21</sup>This number is low relative to other early education experiments. For instance, the student-teacher ratio for the Chicago Child-Parent Center and Expansion Program ranged from 8 to 12 (see Fuerst and Fuerst, 1993).

<sup>22</sup>Schweinhart, Barnes, and Weikart (1993) argue that the certification of the teachers is an important component in the success of the Perry preschool.

<sup>23</sup>The factors used to form the index consist of weighted measures of maternal and paternal education levels, family income, absence of the father from the home, poor social or family support for the mother, indication that older siblings have academic problems, the use of welfare, unskilled employment, low parental IQ, and family members who sought counseling or support from various community agencies. Parental income and education were considered most important in calculating the index.

of the control children were enrolled in preschool and/or kindergarten.

During the first 3 primary school years, a home-school teacher would meet with the parents and help them in providing supplemental educational activities at home. The teacher provided an individually-tailored curriculum for each child. This home-school teacher also served as a liaison between the ordinary teachers and the family, and she would interact with the parents and the teachers about every two weeks. She would also help the family deal with other issues that might improve their ability to care for the child, such as finding employment, navigating the bureaucracy of social services agencies, and transporting children to appointments. Data were collected regularly up to age 21.

### **5.1.3 Chicago Child-Parent Center Program**

The Chicago Child-Parent Center was not evaluated by the method of random assignment. Children in selected neighborhoods were given access to the program and these neighborhoods were matched with comparable areas where the program was not provided. The program was started in 1967 in 11 public schools serving impoverished neighborhoods of Chicago. Using federal funds, the center provided half-day preschool program for 3- and 4-year-olds during the 9 months that they were in school. The program provided an array of services, including health and social services, and free meals. Parental participation was encouraged. Parents were helped to complete school and participated in home visits and field trips. In 1978, state funding became available, and the program was extended through third grade and included a full-day kindergarten experience. Eventually, 24 centers provided preschool and after-school activities, up to second or third grade. This is the period during which the sample analyzed by Reynolds, Ou, and Topitzes (2004) was enrolled in the program. The preschool program ran 3 hours per day during the week for the 9 months that school was in session, and usually included a 6-week summer program. During the kindergarten years, more services were provided at the affiliated school. Teacher-child ratios were 17:2 for the preschool component and 25:2 for the kindergarten. Participation during the primary years was open to any child in the school. Program participants experienced reduced class sizes of 25 pupils rather than the standard of 35 or more in the Chicago public schools. Teachers' aides, extra instructional materials, and enrichment activities were also available. Some children continued to participate in CPC through age 9, for a maximum total of 6 years. 93% of the children were black and 7% were Hispanic.

#### 5.1.4 The Effects of Early Interventions

These and other studies of interventions for children from low-income families find that participants experienced increased achievement test scores, decreased grade retention, decreased time in special education, decreased crime and delinquency and increased high school graduation. The gains vary with quality and age at which the program is started, and there are important differences by the sex of the child.

Programs differ in the measures they use to evaluate the outcomes and in their intensity and quality. As a result, it is hard to compare the programs using a standard basket of benefits. The CPC program, which is less intensive, produced substantial effects on high school graduation rates, reductions in special (remedial) education, grade repetition and juvenile arrest (see Figure 6).

The Perry Preschool Program is the flagship experimental intervention study. Children are followed through age 40. The initial boost in IQ faded by the time the children were in second grade (see Figure 7A), but the program had substantial effects on educational achievement. Achievement test scores for the treatment group were consistently and statistically significantly higher through age 14. Participants had higher grades and were more likely to graduate from high school. Substantially less time was spent in special education, and higher high school graduation rates were achieved by participants (Figure 7B). Participants were more likely to be employed<sup>24</sup> and to earn more (Figure 7C) and they were less dependent on welfare. There was substantially less crime among participants (Figure 7D)—both in terms of incidence and severity, a recurrent finding of early intervention programs. However, there was no statistically significant difference in grade retention by age 27 between the two groups, although teenage pregnancy was lower, and marriage rates were higher by age 27 for program participants.

The Abecedarian program boosted IQ, but its effect is concentrated primarily among girls. Figure 8A shows the overall IQ gap between treatments and controls. It is persistent over time.<sup>25</sup> The Abecedarian program intervenes in the very early years, and it is known that IQ is malleable when children are very young (see, e.g. the discussion in Armor, 2003). This interpretation is reinforced by the observation that the IQ boost was not found among children who only experienced the later intervention. Comparable effects are found for reading scores (Figure 8B) and math achievement scores (Figure 8C). The test score

---

<sup>24</sup>The difference in employment rates was only significant at age 19.

<sup>25</sup>The decline in IQ over time for both groups may be a consequence of the “Flynn Effect” (see Flynn, 1987). Scores are normed against national averages, but over cohorts IQ is increasing.

effects persist through age 21, which is the last age analyzed in the reports available to us.

There were substantial academic benefits as recorded in Figure 8D. Treatment group members participated less in remedial special education at age 15 and repeated fewer grades at all ages. High school graduation and four-year college participation rates were high. Participants were less likely to smoke and had better jobs (see Figure 8E).

Table 1 presents estimated costs and benefits of the Perry and Chicago programs with benefits discounted at a 3% rate. All figures are in 2004 dollars. The benefits vary among programs.<sup>26</sup> Perry produced some gain to parents in terms of reduced child care costs, and earnings gains for participants were substantial. The K-12 benefit arises from the increment in student quality and a reduction in special education costs. This benefit is substantial across all programs. The college/adult category represents the extra tuition paid by students who go to college. Crime represents the reduction in direct costs (incarceration and criminal justice system) as well as damage done to victims. This excludes transfers. Welfare effects are modest. Future Generation (FG) Earnings represents the improvement in the earnings of the descendants of the program participants.

Smoking and health benefits were not measured in the Perry and Chicago studies. For the Abecedarian program, there were substantial effects, including major differences in smoking rates. CPC documents a decline in child abuse and the costs of treating abused children. The costs of Perry were substantial but per year were about the average cost of expenditure on public school students. CPC per year costs about \$6,796 for the preschool and \$3,428 for the school-age component (in 2004 dollars). The benefit cost ratios are substantial: 9 to 1 for Perry; 8 to 1 for Chicago CPC. By projecting from the age 27 results, Rohnick and Grunewald (2003) estimate that the annual rate of return for Perry is 4% for participants and 12% for society at large. Belfield, Nores, and Barnett (2004) use the data on Perry participants through age 40 to estimate that the rate of return for the participants and the general public as a whole is 18.4%. The rate of return varies by the sex of the participants: the rate of return for males alone is 21.9%, while for the rate for females is only 12.6%.<sup>27</sup>

---

<sup>26</sup>There is a cost benefit study of the Abecedarian program (Barnett and Masse, 2002), but it is highly speculative, so that we did not include it here.

<sup>27</sup>Excluding the benefits of the program for the participants, the rate for the general public alone is 16.9%. Belfield, Nores, and Barnett (2004) do not calculate a rate of return for participants only because they do not bear any significant costs of the program. The rate for the general public on investing in males and females separately is 21.0% and 7.6%, respectively. The greater return for men comes from the effect of the intervention on crime, a predominantly male activity.

Some home visitation programs for low-income young mothers have been shown to have modest effects on maternal and offspring behavior and health.<sup>28</sup> Olds (2002) summarizes the results from two randomized trials in Elmira, NY and Memphis, TN, which served predominantly rural white and urban black populations, respectively. The treatment in both trials involved a series of pre- and postnatal home visits of poor, unmarried, and young women by specially-trained nurses.<sup>29</sup> The visits typically lasted 75-90 minutes, and nurses spent more time with women they deemed to have higher needs. The target areas for this intervention were health related behavior during and after pregnancy, childcare skills, and personal development (family planning, education, job search assistance).

The Elmira treatment group made better use of community services and exhibited reduced prenatal-period smoking, with 75% fewer premature deliveries among smokers. At ages 3-4, children whose mothers smoked 10 or more cigarettes during pregnancy had a mean IQ of 4.5 points lower than women who smoked 0-9 cigarettes. Among the 14- to 16-year-old treatment women, the newborn children were almost 400 grams heavier relative to the children of the control women. The beneficial effects of the program were especially apparent for the most disadvantaged women (i.e., young, poor, and unmarried).<sup>30</sup> After the birth of the child, the disadvantaged mothers who were visited showed better parenting skills and higher quality of the home environment. They also had 80% fewer verified cases of child abuse and neglect. Children of visited mothers had 32% fewer visits to the emergency room, and this effect persisted after the end of the program, though the differences in abuse and neglect faded.<sup>31</sup> The disadvantaged subsample of the treatment group had fewer subsequent pregnancies, longer periods between births, and greater employment rates. These effects were also evident by the time the child was 15. The children of the disadvantaged women reported fewer instances of running away, less criminal activity, promiscuous sexual behavior and smoking. Both parents and children reported less use of drugs and alcohol. Importantly, there were no differences in other behavioral problems. A cost-benefit analysis of the Elmira trial by Karoly, Greenwood,

---

<sup>28</sup>Gomby, Larson, Lewit, and Behrman (1999) and Brooks-Gunn, Berlin, and Fuligni (2000) show much more modest effects of home visitation programs, though these implementations are considerably less intensive.

<sup>29</sup>Only women who were pregnant with their first child were eligible. The mean frequency of nurse visits in the prenatal and postnatal (age 0-2) stages were 9 and 23 for Elmira, and 7 and 27 for Memphis. The treatment group was divided into two subgroups, where the first received only prenatal visits. The control group was also divided. See Olds (2002) for more details on the intervention.

<sup>30</sup>This result is found in many studies. Brooks-Gunn, Gross, Kraemer, Spiker, and Shapiro (1992), Magnuson, Meyers, Ruhm, and Waldfogel (2004); Magnuson, Ruhm, and Waldfogel (2004), and Gormley, Gayer, Phillips, and Dawson (2004) find higher effects of the interventions in the disadvantaged population.

<sup>31</sup>This may have been due to improved reporting of abuse by the nurses.

Everingham, Hoube, Kilburn, Rydell, Sanders, and Chiesa (1998) suggests that the program was very successful for low-income, unmarried women. Extrapolating from the results at age 15, the benefits of the program were 4 times its costs. The program paid for itself before the child's fourth birthday, with the primary savings coming from reduced welfare and criminal justice expenditures, as well as increases in tax revenue. However, the program provided no net savings for the sample as a whole, suggesting that targeting, rather than universal provision, is appropriate.

The effects for the Memphis trial were considerably weaker, even for the disadvantaged subsample. There were no effects on birth outcomes and parenting skills. Many fewer women smoked in this sample, so any reductions were very small. The same may be the case for child abuse and neglect. Children of visited women had fewer health-care visits, especially among the disadvantaged subsample. In the first 2 years of life, more visited mothers attempted breast feeding. At age 4, there were no differences in mental development or reported behavior problems. Visited mothers reported fewer subsequent pregnancies. There were no differences in employment and some evidence of reduced AFDC and Food Stamp use. The children are still too young to perform a reliable cost-benefit analysis on their outcomes.

Much more research is needed on Perry, CPC, and the other early childhood program results (shown in Tables 2 and 3). These samples and measurements need to be placed in a common analytical framework to better understand the differences in samples, treatments, and effects. For example, are the persistent Abecedarian effects on IQ due to the intensity or the age (4 months) at which the intervention is administered? How important are home visitation efforts? Joint analysis of the multiplicity of generally favorable treatment outcomes using methods appropriate for the small samples that are available, needs to be applied to supplement analyses of one-at-a-time outcome measure studies. A much more careful analysis of the effects of scaling up the model programs to the target population, and its effects on costs, has to be undertaken before these estimates can be considered definitive.

### **5.1.5 Extreme Deprivation and Remediation**

Institutional rearing of children, insofar as it tends to be exceptionally poor, provides scientists with a unique natural experiment that can be used to ascertain the effects of severe environmental deprivation. Evidence on children from such environments allows us to answer questions about the developmental consequences of negative early experience and how amenable exposed children are to interventions such

as foster care. It may also enable us to learn if there are critical or sensitive periods for development, which would have important implications for the relationship between the timing of an intervention and the extent of its success. Some good evidence on this issue comes from the longitudinal studies of initially institutionalized Romanian infants and toddlers who were later placed into foster care abroad. In this section, we will outline the historical context for these studies, some of their results, and the implications that these data have for our model of human development.

The Ceaușescu regime in Romania, which was in power from 1966 to 1989, attempted to enlarge the country's workforce by increasing the birth rate.<sup>32</sup> Virtually all types of abortion were criminalized, and divorce was made much more difficult. Contraceptives were neither manufactured domestically nor imported. Progressive income taxes on childless adults over 25 were imposed. Monthly cash subsidies were awarded to families with children, and the average allowance per child rose as family size increased. Various labor laws eased working conditions for pregnant and nursing mothers by eliminating overtime and night work entirely, and by reducing physically demanding work. Over three months of paid maternity leave was available, as were additional breaks or reductions in work hours of up to two hours per day. Early retirement was available for women as a function of the number of children they raised to age 10. Increasing economic hardship coupled with Ceaușescu's goal of paying off all international debt by imposing rationing, obliged many women to work outside the home. Since childcare for the young (or any other alternative) was scarce, many children were simply abandoned.

Institutionalization of children was not stigmatized, and was even encouraged officially. When the Ceaușescu regime fell in 1989, there were roughly 170,000 children in 700 overcrowded state institutions (see Rosapepe, 2001). While no rigorous statistics on the conditions in these homes are available, foreign visitors described the situation as appalling (see Rosapepe, 2001; Rutter and the English and Romanian Adoptees Study Team, 1998). Children remained in their cots all day, with no toys or other types of stimulation. Caregiving and personalized affection were all but nonexistent. Many young children were fed only gruel from bottles that were propped up, and some continued to have difficulty even chewing solid food some years later. Orphanages were frequently located in remote areas of the country; some children were transferred far away from where they were born and were "lost" in the system. By the late 1980s, many institutions had no hot water, no constant heat during winter, no diapers or even detergent. Medical

---

<sup>32</sup>Moskoff (1980) enumerates the regime's pronatalist policies.

supplies, including antibiotics and syringe needles, were extremely scarce. Children were often tied down or locked in rooms to keep them under control and some were abused. While the prevalence and incidence of these problems are unknown, most children exhibited a range of emotional, behavioral and medical problems when they were adopted abroad.

Several studies have been conducted to evaluate the effects of interventions at various ages on these children. The largest study of this sort was completed in the UK by Michael Rutter, his colleagues and the English and Romanian Adoptees Study Team. The most recent results are summarized in O'Connor, Rutter, Beckett, Keaveney, Kreppner, the English, and Romanian Adoptees Study Team (2000). This group studied 165 children who were adopted from Romania into UK families between 1990 and 1992 and compared them at ages 4 and 6 to 52 adopted children from within the UK who were all placed before age 6 months.<sup>33</sup> Selected results are shown in Table 4. Rutter and the English and Romanian Adoptees Study Team (1998) shows that at the time of adoption, the orphans showed substantial developmental retardation, malnutrition, and a range of health problems. Relative to ordinary English children, half of the Romanian orphans were below the third percentile on weight, and over a third were below the third percentile on height. The overall mean score on the Denver developmental quotient was 63, indicating mild retardation.<sup>34</sup> Interestingly, there were no significant differences in weight or Denver scale by age of adoption. By age 4, only 2% of the orphans were below the third percentile on weight, and only 1% was below that threshold on height. The extent of catch-up to British adoptees on the Denver developmental quotient was greater for the orphans who entered foster care before they were 6 months of age.<sup>35</sup> At age 6, the same result was obtained.<sup>36</sup> The same pattern appears to hold for cognitive development at ages 4

---

<sup>33</sup>Only 87% of the Romanian children were adopted from institutions. The others came from a family setting, but there were no differences in origin by age at the time of adoption. It is true, however, that the non-institutionalized children exhibited fewer problems.

<sup>34</sup>The Denver Developmental Scales were used to conduct this assessment. Parents were asked to recall specific behavior (e.g. standing while holding on to something, lifting the head, making meaningful "da-da" sounds) at the time of adoption. The majority of parents used baby books that recorded these developmental milestones, which made recollection much better. See Rutter and the English and Romanian Adoptees Study Team (1998) for more details on the analysis.

<sup>35</sup>The mean Denver scale for within-UK adoptees was 117.7 ( $SD = 24.3$ ), 115.7 ( $SD = 23.4$ ) for Romanians adopted before 6 months, and 96.7 ( $SD = 21.3$ ) for those adopted when they were between 6 and 24 months of age. See Rutter and the English and Romanian Adoptees Study Team (1998).

<sup>36</sup>O'Connor, Rutter, Beckett, Keaveney, Kreppner, the English, and Romanian Adoptees Study Team (2000) add a third group of Romanian children who were adopted between the ages of 24 to 42 months. This group exhibits the worst performance on the Denver scale. Due to ceiling effects, the Denver scale is not meaningful at age 6, so O'Connor, Rutter, Beckett, Keaveney, Kreppner, the English, and Romanian Adoptees Study Team (2000) use the presence of impairment (defined as a score below 70) as a test criterion. For within-UK adoptees, only 2% ( $SD = 1$ ) qualify as impaired. The corresponding percentages for the Romanians adopted before 6, 6 – 24 and 24 – 42 months are 0 ( $SD = 0$ ), 5 ( $SD = 2$ ), and 18 ( $SD = 7$ ). See O'Connor, Rutter, Beckett, Keaveney, Kreppner, the English, and Romanian Adoptees Study Team (2000).

and 6, as measured using the McCarthy General Cognitive Index.<sup>37</sup>

Romanian orphans who were adopted into UK families from an environment of severe early deprivation exhibited remarkable improvement. This recovery was characterized by a negative linear dose-response relationship with the duration (or perhaps severity) of the exposure to poor pre and postnatal environments. The children who caught up to ordinary UK adoptees were the ones who were adopted before 6 months of age. This shows the importance of early vs. late intervention that we have documented throughout this paper. This evidence is also consistent with the notion that early environments are a sensitive, rather than a critical period of development for many child outcomes. Had the interventions occurred later in the life of the children, it is likely that they would have been less effective.

## 5.2 Intervention in the Adolescent Years

How effective are interventions in the adolescent years? Is it possible to remedy the consequences of neglect in the early years? These questions are relevant because cognitive abilities are fairly well determined and stable by age 10 in the sense that IQ at later ages is highly correlated with IQ at ages 8-10. Just as early intervention programs have a high payoff primarily from the social skills and motivation they impart to the child and the improved home environment they produce, so do interventions that operate during the adolescent years.

Tables 5 and 6 summarize evidence on the effects of adolescent interventions on education, earnings, and crime rates. There are few estimates of rates of return for these programs. School-based and training-based programs are compared in the table. We briefly discuss what is known about school-based interventions during the adolescent years. A few recent studies of mentoring programs like Big Brothers/Big Sisters (BB/BS) and Philadelphia Futures Sponsor-A-Scholar (SAS) have shown that these programs have broad positive social and academic impacts on participating school-aged children and adolescents. The BB/BS program pairs unrelated adult volunteers with youth from single-parent households for the purpose of providing youth with an adult friend. This activity promotes private youth development and surrogate parenthood. No specific attempts were made to ameliorate particular deficiencies or to reach specific educational goals. A broad, supportive role is envisioned for the mentor.

---

<sup>37</sup>The GCI is the total score on the McCarthy Scales of Children's Abilities. It summarizes verbal, quantitative, perceptual, and memory performance.

In a random-assignment study, Tierney, Grossman, and Resch (1995) found that eighteen months after being matched with a mentor, Little Brothers and Sisters (ages 10 to 16 at the time of the match) were less likely to have initiated drug or alcohol use, to hit someone, to skip class or a day of school, or to lie to their parents; they had higher average grades and were more likely to feel competent in their school work and report a better relationship with their parents.

The primary goal of Sponsor-A-Scholar (SAS) was to help students from Philadelphia public high schools make it to college. The program provides long-term mentoring (throughout high school and for one year beyond), substantial academic support, help with college application and financial-aid procedures, and financial support for college-related expenses. Individually matched mentors served as surrogate parents, provided a successful role model, monitored student progress, and provided encouragement and support. SAS provided students with \$6,000 in financial assistance throughout college for those choosing to enroll in an accredited two- or four-year postsecondary institution. The program also provided a coordinator for groups of about thirty students to ensure a successful relationship is built between mentors and students. Using a matched sample of non-SAS students in Philadelphia high schools, Johnson (1996) estimates statistically significant increases in grade point averages for tenth and eleventh grades, as well as a 22 percent (16 percent) increase in college attendance one year (two years) after graduation from high school. Because the primary goal of SAS is to increase college enrollment, Johnson did not collect other social and psychological measures.

Much like SAS, the Quantum Opportunity Program (QOP) offered disadvantaged minority students counseling and financial incentives (one dollar up front and one dollar put in a college fund) for every hour spent in activities aimed at improving social and market skills. Students who were randomly chosen to participate in the program were provided with a mentor at the beginning of ninth grade. All participants were kept in the program for four years regardless of whether they stayed in school. Over four years, the average participant logged 1,286 hours of educational activities like studying with tutors or visiting museums. Two years after program completion, about a third more participating students graduated from high school (or obtained a GED) than similar nonparticipants. Since many participants were enrolled in postsecondary schooling at the time of the follow-up study, it is difficult to determine the program's effect on earnings. Arrest rates for program participants, however, were one-half those for nonparticipants. These benefits did not come without substantial expenditures, however, as the average four-year cost per

participant was \$10,600. Still, a cost-benefit analysis estimated positive net social returns to QOP. (See Taggart, 1995, for a more detailed description of the program and an evaluation of its impacts). Tables 5 and 6 present evidence from a randomized-trial evaluation of the QOP program. Again, the evidence shows that QOP and programs like it can dramatically improve social skills and the adaptation of adolescents to society. However, these programs do not produce miracles. The recent evaluation of QOP by Maxfield, Schirm, and Rodriguez-Planas (2003) found that the program did not improve grades or achievement test scores and the effect on risky behaviors was ambiguous. It was also more effective for teens from the middle of the eligible grade distribution than for enrollees at the top or bottom of the distribution. There was considerable variability in estimated program effects by site.

Two other studies provide additional evidence that creative programs designed to keep adolescents in school can be effective. These are discussed more extensively in Heckman (2000) and Heckman and Lochner (2000), and we briefly summarize these discussions here. Ohio's Learning, Earning, and Parenting (LEAP) program and the Teenage Parent Demonstration (TPD) provided financial incentives for teenage parents on welfare to stay in school or take GED classes (or, alternatively, imposed financial penalties for nonenrollment). LEAP showed increases in high school graduation or GED rates among randomly assigned participants who were still enrolled in school when they entered the program. TPD showed mixed results on educational attainment depending on the program site. Young women who had already dropped out of school at the time of enrollment in the program (and, to a lesser extent, those who were still attending school when they entered the program) may have substituted GED training for high school graduation as an easier way to meet program requirements, raising concerns about an unintended, potentially negative effect. Both of these programs show positive post-program effects on earnings and employment for students who were still in school when they entered the program. The estimated effects were often negative, however, for participants who had already dropped out of school before entering the program. Both studies thus show more positive impacts for individuals still enrolled in school than for dropouts. It is still unknown whether the effects of the programs are more positive for those still in school because, on average, they are of higher ability than those who have already dropped out, or because there is some advantage to intervening before adolescents leave school.

The available schooling literature demonstrates that providing disadvantaged students with financial incentives to stay in school and participate in learning activities can increase schooling and improve em-

ployment outcomes. It should be noted that although programs providing such incentives have proven to influence employment and earnings positively (and, in the case of QOP, to reduce crime), they do not perform miracles. The impacts they achieve are modest, but positive.

The Summer Training and Employment Program (STEP) provided remedial academic education and summer jobs to disadvantaged youth ages 14 and 15. Each summer, participants enrolled in 110 hours of classes and 90 hours of part-time work. Although program participants achieved modest short-term gains in reading and math skills, those gains did not last. Two to three years after program completion, program participation was found to have no effects on high school graduation rates, grades, or employment (see Table 10). The program has been criticized for not attempting to follow up on its summer program with a school year curriculum. Maryland's Tomorrow program did just that: it combined an intensive summer program with a school year follow-up, offering participants summer jobs and academic instruction, career guidance, and counseling through adult mentors, peer support, or tutoring. Although the program did not reduce final attrition rates, it did seem to delay attrition (dropout rates were lower for program participants during the ninth grade but not by the end of the twelfth grade). The program also increased the pass rate for twelfth grade students taking the Maryland Functional Tests, a series of tests of basic skills (see Heckman and Lochner, 2000).

There is also some non-experimental evidence that Catholic secondary schooling is associated with increased college participation among urban students, especially minorities (see Grogger and Neal, 2000). This increase does not appear to be accompanied by large gains in math scores, at least for the groups whose attainment is most affected. This is consistent with our hypothesis that adolescent interventions alter noncognitive skills but have weaker effects on cognitive skills. Altonji, Elder, and Taber (2005) find a similar pattern that attendance at Catholic schools raises high school graduation rates and, more tentatively, promotes college attendance but has no effect on test scores.

The evidence on programs aimed at increasing the skills and earnings of disadvantaged youth suggests that sustained interventions targeted at adolescents still enrolled in school can positively affect learning and subsequent employment and earnings. The studies discussed in this Section also suggest that interventions for dropouts are much less successful. One plausible interpretation, consistent with other evidence reported in this paper, is that those who choose to drop out have less motivation and lower ability, making programs less effective for them regardless of when the intervention takes place. It is important to note, however,

that the interventions conducted by such programs only alleviate and do not fully reverse early damage caused by low quality family environments.

### **5.3 The Effectiveness of Late Adolescent and Young Adult Remediation Programs**

The evidence from public job training and second chance programs like the GED suggests that remediation targeted towards children from disadvantaged environments is costly and at current expenditure levels is ineffective. Heckman, LaLonde, and Smith (1999) survey evaluations of public job training programs in the United States. Returns are low (and sometimes negative) and even when they are positive they do not lift most persons treated out of poverty. Similar evidence is reported for remediation efforts in public schools. As we discussed above, the return to GED certification is very low. While the return to private sector on-the-job training is high, access to such training is difficult for the less able and the disadvantaged. Adolescent remediation programs are effective for a targeted few who use them as second chance opportunities. They are not effective for the rest.

Some look to public schooling as a way to remedy early ability deficits and to alleviate disadvantage in endowments. Hansen, Heckman, and Mullen (2004) and Heckman, Larenas, and Urzua (2004) address this issue. A variety of methods show that schooling, while it raises measured ability, does not eliminate gaps between children from different racial and economic strata, and if anything widens them. Experience raises performance but does not close gaps.

Figures 9A-B, taken from Heckman, Larenas, and Urzua (2004), show how schooling raises achievement test scores at different levels of ability (AFQT is a measure of achievement). These authors use the methodology of Hansen, Heckman, and Mullen (2004) to isolate causal effects of schooling on AFQT test scores, holding cognitive ability constant. Their analysis is based on longitudinal data to measure the effects of different levels of schooling attained at the date the test is taken on achievement for people who all eventually get the same schooling. For all major demographic groups, initial (ninth grade) test score gaps are maintained regardless of schooling level. Schooling raises test scores, but it does not equalize them. These results persist even after controlling for measures of schooling quality. One cannot count on schooling to eliminate early test score deficits. On the other hand, one cannot blame schools for widening

initial test score gaps.

The evidence reviewed in this section points to the empirical importance of self-productivity and complementarity. Skill begets skill. Later remediation of early skill deficits can be costly. We next present a more formal model of the technology of skill formation that is a starting point for the theoretical unification of a scattered literature on treatment effects that presents “effects” for different programs in different environments directed towards different clientele.

## Part III

# Using the Technology of Skill Formation to Explain the Evidence

## 6 A Model of Skill Formation

We use simple economic models to organize the evidence presented in Part II. We define the concepts of recursive productivity or “self-productivity” and complementarity and show how the skill multiplier (as defined in this section) and the notion of complementarity help to organize the empirical evidence surveyed in Part II. These concepts are essential for understanding why early interventions are more effective than later interventions and why there is no trade-off between equity and efficiency in the early years of childhood but why there is such a trade-off in the later years.

In the models presented in this section, parents make decisions about their children. We ignore how the parents get to be who they are and the decisions of the children about their own children. See Cunha and Heckman (2006) for a model of intergenerational transmission of skill.

Suppose that there are  $T$  periods in a child’s life before the child becomes an adult. Adulthood starts at period  $T + 1$ . The child works for a fixed number of periods after the  $T$  periods of childhood. Models based on the analysis of Becker and Tomes (1979) assume only one period of childhood. To fix ideas, we consider first the case in which there is one type of skill. We relax this assumption below and explicitly

model cognitive and noncognitive skills. Furthermore, we provide evidence that the accumulation processes over childhood and adolescence are different for these two distinct skills.

Let  $I_t$  denote parental investments in child skill at period  $t$  and  $t = 1, 2, \dots, T$ . The parents fully control the investments in the skills of the child. We first describe how skills evolve over time. Assume that each agent is born with initial conditions  $\theta_1$ . At each stage  $t$  let  $\theta_t$  denote the vector of skill stocks. Let  $\theta_{T+1}$  denote the level of skills as the child starts adulthood. The technology of production of skill at period  $t$  is

$$\theta_{t+1} = f_t(\theta_t, I_t). \quad (1)$$

for  $t = 1, 2, \dots, T$ . We assume that  $f_t$  is twice continuously differentiable, increasing and concave in  $I_t$ .

Note that the technology (1) is recursive. One can rewrite the stock of skills at stage  $t + 1$ ,  $\theta_{t+1}$ , as a combination of past investments:<sup>38</sup>

$$\theta_{t+1} = m_t(\theta_1, I_1, \dots, I_t). \quad (2)$$

Technology (1) and its non-recursive representation (2) are sufficiently rich to describe learning in rodents and rhesus monkeys as documented by Meaney (2001) and Cameron (2004). It also captures the critical and sensitive periods in animals documented by Knudsen, Heckman, Cameron, and Shonkoff (2006). Emotionally nurturing early environments create preconditions for later cognitive learning. More emotionally secure young animals explore their environments more actively and learn more quickly. This is an instance of complementarity.

Period  $t^*$  is a *critical period* for  $\theta_t$  if

$$\frac{\partial \theta_{t+1}}{\partial I_s} = \frac{\partial m_t(\theta_1, I_1, \dots, I_t)}{\partial I_s} \equiv 0 \text{ for all } \theta_1, I_1, \dots, I_t, s \neq t^*,$$

but

$$\frac{\partial \theta_{t+1}}{\partial I_{t^*}} = \frac{\partial m_t(\theta_1, I_1, \dots, I_t)}{\partial I_{t^*}} > 0 \text{ for some } \theta_1, I_1, \dots, I_t.$$

---

<sup>38</sup>For  $t + 1 = 2$  note that:

$$m_1(\theta_1, I_1) = f_1(\theta_1, I_1)$$

For  $t + 1 = 3$  then:

$$m_2(\theta_1, I_1, I_2) = f_2(f_1(\theta_1, I_1), I_2).$$

For other stages  $t + 1 > 3$  we perform a recursion on functions  $f_t$  to derive the appropriate function  $m_t(\cdot)$ .

This says that investments in  $\theta_{t+1}$  are productive in period  $t^*$  but not in any other period  $s \neq t^*$ . There are critical periods for the development of vision in humans. Cataracts in babies and young children are treated urgently because they can have a lasting effect on vision development. As the cloudy lens blocks light from getting into the eye, the brain gets no visual experience through that eye at a time when the eye and brain are working together to learn to see. As a result, a baby or child with an untreated cataract could be slowly going blind, and if surgery is delayed, it might be too late to help.

Period  $t^*$  is a *sensitive period* for  $\theta_{t+1}$  if

$$\left. \frac{\partial \theta_{t+1}}{\partial I_s} \right|_{\theta_1=\theta, I_1=i_1, \dots, I_t=i_t} < \left. \frac{\partial \theta_{t+1}}{\partial I_{t^*}} \right|_{\theta_1=\theta, I_1=i_1, \dots, I_t=i_t}.$$

Thus period  $t^*$  is a sensitive period if, at the same level of inputs, investment is more productive in stage  $t^*$  than in other stage  $s \neq t^*$ <sup>39</sup>. The evidence in Part II suggests that there are sensitive periods for the acquisition of some parts of language such as learning how to speak a foreign language without accent.

Suppose that  $T = 2$ . The adult stock of skills,  $\theta_3$ , is a function of initial conditions and investments during childhood:

$$\theta_3 = m_2(\theta_1, I_1, I_2). \quad (3)$$

The literature in economics assumes only one period of childhood. It does not distinguish between investments in early from late periods. In its simplest formulation, this can be represented by the special case:

$$\theta_3 = m_2(\theta_1, \gamma I_1 + (1 - \gamma) I_2), \quad (4)$$

with  $\gamma = \frac{1}{2}$ . Equation (4) states that adult stocks of skills do not depend on how investments are distributed over different periods of childhood. For example, take two children,  $A$  and  $B$ , such that they have the same initial condition,  $\theta_1^A = \theta_1^B = \theta_1$ , but they have different investment profiles: child  $A$  receives no investments in period one and receives  $I$  units of investment in period two,  $I_1^A = 0$ ,  $I_2^A = I$ , while child  $B$  receives  $I$  units of investment in period one and zero units of investment in period two,  $I_1^B = I$ ,  $I_2^B = 0$ . According to (4) children  $A$  and  $B$  will have the same stocks of skills as adults. The predictions from technology (4) are a little extreme and certainly at odds with the empirical evidence summarized in Part II.

---

<sup>39</sup>See Cunha, Heckman, Lochner, and Masterov (2006) for a definition of critical and sensitive periods in terms of the technology (1).

One can illustrate the importance of the timing of investments by considering another extreme example:

$$\theta_3 = m_2(\theta_1, \min\{I_1, I_2\}). \quad (5)$$

Equation (5) states that adult stocks of skills depend on how investments are distributed over time. For example, if investments in period one are zero,  $I_1 = 0$ , then it does not pay to invest at period two, because  $\min\{I_1, I_2\} = \min\{0, I_2\} = 0$  for any  $I_2 > 0$ . For the same reason, if late investments are zero,  $I_2 = 0$ , it does not pay to invest early. As we will show below, for the technology of skill formation defined by (5), the best strategy is to distribute investments evenly, so that  $I_1 = I_2$ .

A more general representation that captures both (4) and (5) as special cases is given by:

$$\theta_3 = m_2\left(\theta_1, \left[\gamma (I_1)^\phi + (1 - \gamma) (I_2)^\phi\right]^{\frac{1}{\phi}}\right) \quad (6)$$

for  $\phi \leq 1$  and  $0 \leq \gamma \leq 1$ .

The parameter  $\gamma$  is a *skill multiplier*. It arises because  $I_1$  affects the accumulation of  $\theta_2$  which in turn affect the productivity of  $I_2$  in forming  $\theta_3$ . Thus  $\gamma$  captures the net effect of  $I_1$  on  $\theta_3$  through both self-productivity and direct complementarity.

The number  $\frac{1}{1-\phi}$  is a measure of how easy it is to substitute between  $I_1$  and  $I_2$ . Within the CES technology,  $\phi$  represents the degree of complementarity (or substitutability) between early and late investments in producing skills. In this role, the parameter  $\phi$  dictates how easy it is to compensate for low levels of stage 1 skills in producing late skills.

When  $\phi$  is small, low levels of early investment  $I_1$  are not easily remediated by later investment  $I_2$  in producing human capital. The other face of CES complementarity is that when  $\phi$  is small, high early investments should be followed with high late investments. In the extreme case when  $\phi \rightarrow -\infty$ , (6) converges to (5). As discussed above, the Leontief case contrasts sharply with the case (4), which arises when  $\phi = 1$  and the further restriction that  $\gamma = \frac{1}{2}$ .

In analyzing the optimal timing of investment, it is convenient to work with the technology embodied in (6). We now show how the ratio of early to late investments varies as a function of  $\phi$  and  $\gamma$ . Consider

the following model in which parents maximize the present value of net wealth of their children.<sup>40</sup> In order to do that, parents decide how much to invest in period “1,”  $I_1$ , how much to invest in period “2,”  $I_2$ , and how much to transfer in risk-free assets,  $b$ , given total parental resources  $M$ . Parents cannot extract resources from children, so  $b \geq 0$ . From period “3” to period  $T_R$ , the age of retirement from the workforce, persons are assumed to work full time. Let  $r$  denote the time-invariant interest rate, set exogenously and assumed to be constant for all periods, and let  $q$  denote the present value of future earnings per efficiency unit of human capital  $\{w_t\}_{t=3}^T$ :

$$q = \sum_{t=3}^{T_R} \left( \frac{1}{1+r} \right)^{t-3} w_t.$$

Lifetime earnings of children when they start working at period “3” are given by  $q\theta_3$ . Discounted to period 1, the present value of lifetime earnings is  $\frac{q}{(1+r)^2}\theta_3$ . The problem of the parents is to maximize the present value of the child’s net wealth:

$$\max_{I_1, I_2, b} \left\{ \frac{1}{(1+r)^2} [q\theta_3 + b] \right\},$$

subject to the technology of skill formation (6), the standard budget constraint

$$I_1 + \frac{1}{1+r} I_2 + \frac{1}{(1+r)^2} b = M, \tag{7}$$

and the constraint that parents cannot leave negative bequests to their children

$$b \geq 0, \tag{8}$$

When  $\phi = 1$ , early and late investments are perfect CES substitutes. The optimal investment strategy for this technology in this simple environment is straightforward. The price of early investment is \$1. The price of the late investment is  $\frac{1}{(1+r)}$ . Thus the parents can purchase  $(1+r)$  units of  $I_2$  for every unit of  $I_1$ . The amount of human capital produced from one unit of  $I_1$  is  $\gamma$ , while  $\$(1+r)$  of  $I_2$  produces  $(1+r)(1-\gamma)$  units of human capital. Therefore, the parent invests early if  $\gamma > (1-\gamma)(1+r)$  and late otherwise. Two forces act in opposite directions. High productivity of initial investment (the skill multiplier

---

<sup>40</sup>This setup is overly simplistic but allows us to focus on the important points. See Caucutt and Lochner (2004), Cunha (2004), and Cunha and Heckman (2006) for more general models.

<sup>40</sup>We abstract from endogenously determined on-the-job training, learning-by-doing, and assume that agents supply labor inelastically.

$\gamma$ ) drives the agent toward making early investments. Intertemporal prices (the interest rate) drive the agent to invest late. It is optimal to invest early if  $\gamma > (1 - \gamma)(1 + r)$ .

As  $\phi \rightarrow -\infty$ , the CES production function converges to the Leontief case and the optimal investment strategy is to set  $I_1 = I_2$ . In this extreme case, CES complementarity has a dual face. Investments in the young are essential. At the same time, later investments are needed to harvest early investments. On efficiency grounds, early disadvantages should be perpetuated, and compensatory investments at later ages are economically inefficient.

For  $-\infty < \phi < 1$ , the first-order conditions are necessary and sufficient given concavity of the technology in terms of  $I_1$  and  $I_2$ . Notice that if restriction (8) is not binding, then optimal early and late investments are only functions of  $(q, r)$ . In this case, all unconstrained families that make bequests will invest the same in their children. The only difference is in the transfers of assets to their children. If  $M_A > M_B$  then  $b_A > b_B$ .

For an interior solution we can derive the optimal early to late ratio of investments:

$$\frac{I_1}{I_2} = \left[ \frac{\gamma}{(1 - \gamma)(1 + r)} \right]^{\frac{1}{1 - \phi}}. \quad (9)$$

Figure 10 plots the ratio of early to late investments as a function of the skill multiplier  $\gamma$ , under different values of the complementarity parameter  $\phi$ . When  $\phi \rightarrow -\infty$ , we obtain the Leontief technology and there is high CES-complementarity between early and late investments. In this case, the ratio is not sensitive to variations in  $\gamma$ . CES-complementarity dominates, and the optimal investment profile distributes investments equally across different periods. When  $\phi = 0$ , the function (6) is given by the Cobb-Douglas function:

$$h = m_2 (\theta_1, (I_1)^\gamma (I_2)^{1 - \gamma}).$$

In this case, from equation (9),  $\frac{I_1}{I_2}$  is close to zero for low values of  $\gamma$ , but explodes to infinity as  $\gamma$  approaches one.

The lessons we take from this simple analysis are summarized in Table 7. When CES complementarity is high, the skill multiplier  $\gamma$  plays a limited role in shaping the ratio of early to late investments. High early investments should be followed by high late investments. As the degree of CES complementarity decreases, the role of the skill multiplier increases, and the higher the multiplier, the more investments

should be concentrated in the early ages.

This simple model also has implications for the timing of interventions. Suppose that  $M_A > M_B$  and family  $A$  is unconstrained while family  $B$  is constrained. Consequently, in equilibrium, the marginal return to one dollar invested in the poor child from family  $B$  is above the marginal return to the same dollar invested in the rich child from family  $A$ , so family  $B$  underinvests compared to the unconstrained family  $A$ .

There is no trade-off between equity and efficiency in *early* childhood investments. Government policies to promote early accumulation of human capital should be targeted to the children of poor families. However, the optimal second period intervention for a child from a disadvantaged environment depends critically on the nature of the technology (6). If  $I_1$  and  $I_2$  are perfect CES complements, then a low level of  $I_1$  cannot be compensated at any level of investment by a high  $I_2$ . On the other hand, suppose that  $\phi = 1$ , so the reduced form technology can be written with inputs as perfect CES substitutes (4). Then a second-period intervention can, in principle, eliminate initial skill deficits (low values of  $I_1$ ). At a sufficiently high level of second-period investment, it is technically possible to offset low first period investments. However, it may not be cost effective to do so. For example, if  $q(1 - \gamma) < 1 + r$ , then the gains from future earnings do not justify the costs of investment. It would be more efficient to give the child a bond that earns interest rather than to invest in human capital in order to put the child at a certain level of income. Carneiro and Heckman (2003) show that classroom size reductions at current levels of funding in the U.S. are an example of such a policy.

We previously discussed the concepts of critical and sensitive periods in terms of the technical possibilities of remediation. These were defined in terms of the technology of skill formation. Here, we consider the net effects operating through investment and market substitution. The higher  $\phi$ , the greater are the possibilities for alleviating early disadvantage. When  $\phi = 1$ , as in this example, it is always technically possible to remediate early disadvantage. But it may not be economically efficient to do so. From an economic point of view, critical and sensitive periods should be defined in terms of the costs and returns of remediation, and not solely in terms of technical possibilities.

The model developed above assumes that skills can be represented by a one-dimensional object  $\theta_t$ . The evidence summarized above shows the importance of both cognitive and non-cognitive skills. It is important to understand the accumulation process of each one of these skills. Accordingly, let  $\theta_t$  denote

the vector of cognitive and noncognitive skills:  $\theta_t = (\theta_t^C, \theta_t^N)$ . At each stage  $t$ , we can define a recursive technology for cognitives,

$$\theta_{t+1}^C = f_t^C (\theta_t^C, \theta_t^N, I_t), \quad (10)$$

and another one for noncognitive skills,

$$\theta_{t+1}^N = f_t^N (\theta_t^C, \theta_t^N, I_t). \quad (11)$$

Note that the technologies (10) and (11) allow for cross-productivity effects: cognitive skills may affect the accumulation of noncognitive skills and vice versa.

We can extend the skill formation to allow for parental skills to affect directly the accumulation of skills of children. Let  $\theta_P^C, \theta_P^N$  denote the stocks of cognitive and noncognitive skills of the parents. Then, the technology for the formation of cognitive skill  $\theta_t^C$  can be defined as:

$$\theta_{t+1}^C = f_t^C (\theta_t^C, \theta_t^N, I_t, \theta_P^C, \theta_P^N). \quad (12)$$

Similarly, for noncognitive skills, we would have:

$$\theta_{t+1}^N = f_t^N (\theta_t^C, \theta_t^N, I_t, \theta_P^C, \theta_P^N). \quad (13)$$

We can parameterize the technology functions (12) and (13) according to a CES specification:

$$\theta_{t+1}^k = B_{t,k} \left[ \gamma_{t,k,1} (\theta_t^C)^{\phi_{t,k}} + \gamma_{t,k,2} (\theta_t^N)^{\phi_{t,k}} + \gamma_{t,k,3} (I_t)^{\phi_{t,k}} + \gamma_{t,k,4} (\theta_P^C)^{\phi_{t,k}} + \gamma_{t,k,5} (\theta_P^N)^{\phi_{t,k}} \right]^{\frac{1}{\phi_{t,k}}} \quad (14)$$

and  $\sum_{l=1}^5 \gamma_{t,k,l} = 1$ . One obtains the technology for cognitive skills when the index  $k$  is equal to  $C$ . When the index  $k$  is equal to  $N$  one obtains the technology for noncognitive skills. Using the function (14) we can compare, for example, how easy it is at stage  $t$  to substitute early for late investments in cognitive versus noncognitive skills, which can be done by comparing  $\phi_{C,t}$  against  $\phi_{N,t}$ . We can compare the impact in stage  $t$  of parental investments on the accumulation of cognitive and non-cognitive skills, which is summarized by a comparison between  $\gamma_{t,C,3}$  against  $\gamma_{t,N,3}$ .

These are theoretical models. In the next section we put empirical flesh on the theoretical skeleton.

## Part IV

# Estimating the Technology for the Formation of Skills

## 7 The Estimation of the Technologies for the Formation of Cognitive and Noncognitive Skills

Estimation of the technology (12) and (13) would be straightforward if we observed the dependent variables, which are the next stage cognitive and noncognitive skills,  $\theta_{t+1}^C$ ,  $\theta_{t+1}^N$ , as well as the independent variables which are the current stage stocks of cognitive and noncognitive skills,  $\theta_t^C$ ,  $\theta_t^N$ ; parental investments,  $I_t$ ; and parental stocks of cognitive and noncognitive skills,  $\theta_P^C$ ,  $\theta_P^N$ . Unfortunately, we do not observe any of these measures directly.

One way to solve the problem is to use the scores from cognitive and noncognitive tests and assume that they are perfect measurements of each of these skills. A similar measurement problem arises when it comes to construct parental investments. Sometimes parental investments are not observed directly, and researchers often use family income or parental education as proxies. In both situations, one cannot usually test the validity of these proxies, which are likely to be imperfect at best. Another problem is that we sometimes observe what may be some of the components of parental investments. In this case, the question becomes how these different components should be combined to produce meaningful measures.

The approach developed by Cunha, Heckman, and Schennach (2006) to estimate the full technology functions (12) and (13) recognizes explicitly that the available data are imperfect measurements of the stocks of skills and parental investments. It explores the richness of data sets that usually contain many different test scores designed to measure cognitive and noncognitive skills. Their methodology also allows researchers to construct the optimal combination of the data on the different components to form the parental investment variables.

We provide an intuitive discussion on their methodology. To fix ideas, consider the case in which at

every stage  $t$  of childhood, for each child  $i$  in the dataset, we observe the scores on two different tests that measure cognitive skills,  $Z_{i,t,1}^C, Z_{i,t,2}^C$ ; two scores on tests designed to measure noncognitive skills, which we denote by  $Z_{i,t,1}^N$  and  $Z_{i,t,2}^N$ ; and two components of parental investments,  $Z_{i,t,1}^I, Z_{i,t,2}^I$ . For example, many data sets reports scores on math ( $Z_{i,t,1}^C$ ) and reading ( $Z_{i,t,2}^C$ ). The same data set may report scores in risk-aversion ( $Z_{i,t,1}^N$ ) and patience ( $Z_{i,t,2}^N$ ). Finally,  $Z_{i,t,1}^I$  may be the number of books child  $i$  has at stage  $t$ , while  $Z_{i,t,2}^I$  may be how often the children has tutoring classes. Assume that the same (or similar) observations for parental cognitive and noncognitive skills are also available. Let  $Z_{i,P,1}^C, Z_{i,P,2}^C$  denote the parental scores for cognitive tests. Analogously, we use  $Z_{i,P,1}^N, Z_{i,P,2}^N$  to denote the parental scores in noncognitive tests.

Test scores are noisy measures. For the mean-zero cognitive test score  $Z_{i,t,1}^C$ , let  $\varepsilon_{i,t,1}^C$  denote the difference between the observed test score and the unobserved cognitive skill:

$$Z_{i,t,1}^C = \theta_{i,t}^C + \varepsilon_{i,t,1}^C. \quad (15)$$

Suppose that  $E(\varepsilon_{i,t,1}^C) = 0$  but  $\text{Var}(\varepsilon_{i,t,1}^C) > 0$ . With these assumptions, equation (15) accounts for discrepancies between the observed score  $Z_{i,t,1}^C$  and unobserved skills  $\theta_{i,t}^C$ . However, because of the noise  $\varepsilon_{i,t,1}^C$ ,  $Z_{i,t,1}^C \neq \theta_{i,t}^C$ , so one can not infer skills directly from score.

Assume now that we have a second mean-zero cognitive test score:

$$Z_{i,t,2}^C = \theta_{i,t}^C + \varepsilon_{i,t,2}^C.$$

Can we fix the error in (15)? The answer is yes. Intuitively, with two measures on the same  $\theta_{i,t}^C$ , we can construct an average that reduces the measurement error. Suppose that the noise in one score is independent from the noise in the other score. Suppose that the noises  $\varepsilon_{i,t,1}^C$  and  $\varepsilon_{i,t,2}^C$  are also independent from the unobserved skill  $\theta_{i,t}^C$ . To make the argument simple, suppose that all these variables are normally distributed:  $\varepsilon_{i,t,1}^C \sim N(0, \sigma_{c,t,1}^2)$ ,  $\varepsilon_{i,t,2}^C \sim N(0, \sigma_{c,t,2}^2)$ , and  $\theta_{i,t}^C \sim N(0, \sigma_{\theta_{c,t}}^2)$ . In this case, the distribution of  $\theta_{i,t}^C$  depends only on the variance of unobserved skills  $\sigma_{\theta_{c,t}}^2$  which can be directly computed by the covariance between  $Z_{i,t,1}^C$  and  $Z_{i,t,2}^C$ :

$$\text{Cov}(Z_{i,t,1}^C, Z_{i,t,2}^C) = \sigma_{\theta_{c,t}}^2.$$

We can proceed in the same fashion and obtain the entire joint distribution of unobserved cognitive

skills, unobserved noncognitive skills, unobserved parental investment, unobserved parental cognitive skills, and unobserved parental noncognitive skills. This helps us estimate the technology of skill formation. The analysis of Cunha, Heckman, and Schennach (2006) shows that to estimate the technology functions (12) and (13) all that we need is to know the joint distribution of the skills, and not the amount of skill of each person  $i$ .

This can be made clearer with the following example. Suppose that we want to estimate the coefficient  $\beta$  in the following simple linear regression relationship between the scalar variables:

$$Y = X\beta + \varepsilon,$$

where  $X$  is independent from  $\varepsilon$ ,  $E(\varepsilon) = 0$ ,  $\text{Var}(\varepsilon) = \sigma^2 I$ . Assume that we do not observe either  $Y$  or  $X$ , but for some reason we know the joint distribution  $F(Y, X)$ . Then, we can immediately construct the moments  $E(YX)$  and  $E(X^2) \neq 0$ . We can estimate the parameter  $\beta$  by least squares:

$$\hat{\beta} = \frac{E(YX)}{E(X^2)}.$$

To obtain such estimator, we don't need to observe the data if we know the moments. The same principle applies to the estimation of the technology of skill formation. Given knowledge of the joint distributions, we can construct the appropriate moments for the estimation of the parameters in the functions (12) and (13).

Another advantage of the approach of Cunha, Heckman, and Schennach (2006) is that it can also be used to infer how informative each of the test scores really are on unobserved skills. Note that the variance of the test score can be decomposed as

$$\text{Var}(Z_{i,t,1}^C) = \text{Var}(\theta_{i,t}^C) + \text{Var}(\varepsilon_{i,t,1}^C).$$

If, for example,  $\frac{\text{Var}(\theta_{i,t}^C)}{\text{Var}(Z_{i,t,1}^C)} = 1$ , then the test score  $Z_{i,t,1}^C$  is a perfect measure of the skills  $\theta_{i,t}^C$ . If, on the other hand,  $\frac{\text{Var}(\theta_{i,t}^C)}{\text{Var}(Z_{i,t,1}^C)} = 0$ , then the test score  $Z_{i,t,1}^C$  is not informative about the skills  $\theta_{i,t}^C$  at all. It is pure noise.

The same principle can be used to construct parental investments from observed data. For example,

consider the number of books available to the child. This variable is correlated with parental inputs because parents who invest more in the development of their children will tend to spend more resources on books. But the number of books is unlikely to be a perfect indicator of total parental input. Suppose that the number of books child  $i$  has at age  $t$  ( $Z_{i,t,1}^I$ ) can be written as

$$Z_{i,t,1}^I = I_{i,t} + \varepsilon_{i,t,1}^I.$$

Assume we observe the number of words the mother reads to the child,  $Z_{i,t,2}^I$  and how often the child is taken to a museum,  $Z_{i,t,3}^I$ . We model these variables as:

$$Z_{i,t,2}^I = \alpha_{t,2}^I I_{i,t} + \varepsilon_{i,t,2}^I,$$

and

$$Z_{i,t,3}^I = \alpha_{t,3}^I I_{i,t} + \varepsilon_{i,t,3}^I.$$

The coefficients  $\alpha_{t,2}^I$  and  $\alpha_{t,3}^I$  are called factor loadings. Assuming the errors  $\varepsilon_{i,t,k}^I$ ,  $k = 1, 2, 3$ , are mutually independent and independent of  $I_{i,t}$ , the factor loading can be estimated along with the technology functions and are fundamental information for construction of the parental investment variable.

One can interpret the inverse of the factor loadings on the investment inputs as a measure of the strength of the relationship between the measure  $Z_{i,t,k}^I$  and the unobserved parental investment variable,  $I_t$ . For every measurement  $Z_{i,t,k}^I$  we obtain the relationship:

$$\frac{1}{\alpha_{t,k}^I} E(Z_{i,t,k}^I) = E(I_{i,t}).$$

We can construct the implicit relative weights on the inputs in predicting  $I_t$ ,  $w_{t,k}$ ,  $k = 1, 2, 3$ :

$$w_{t,k} = \frac{\frac{1}{\alpha_{t,k}^I}}{\frac{1}{\alpha_{t,1}^I} + \frac{1}{\alpha_{t,2}^I} + \frac{1}{\alpha_{t,3}^I}}.$$

This is a weighted average of the observed inputs that proxies  $I_t$ , which is a measure of the true parental investment variable.

Noisy measurement of variables is only one of the problems entailed in estimating technology functions (12) and (13) from test score data. Another problem is that test scores do not have a natural metric. Consider, for example, the Stanford-Binet test scores. The scores are scaled such that, for each age group, the population distribution of scores has mean equal to 100 and standard deviation equal to 16. Let the variable  $X$  denote a score measured on the scale of the Stanford-Binet. Consider now a transformation of that score along the following lines. Let  $X_{k,\min}$ ,  $X_{k,\max}$  denote the minimum and maximum score for age group  $k$  in the original Stanford-Binet scale. For each score  $X$  in the group age  $k$ , generate the score  $Y$  such that it satisfies:

$$Y = 10 \left( \frac{X - X_{k,\min}}{X_{k,\max} - X_{k,\min}} \right).$$

The scores  $Y$  are scaled in a way that for each age group  $k$  the lowest score is always zero and the highest score is always a perfect ten. There is no loss information in going from reporting scores in the original scale  $X$  to the alternative scale  $Y$ . Both are valid. There are many other valid ways of reporting the scores.

This richness in choice is troublesome for using test scores as a measure of output. The first question that arises is: Do different metrics imply different parameter values for the technology? In general we would expect so, especially if the technology functions are nonlinear, as are the technologies (12) and (13). The second question is: if different metrics generate different results, and if all the different metrics are equally valid, then which one should the analyst pick?

There is no satisfactory answer to this question looking solely at test scores. If test scores do not have a natural metric, the solution is to find an outcome (other than test scores) that (a) does have a natural metric and (b) are correlated with cognitive and noncognitive skills. Consider the logarithm of monthly wages when the person is thirty years-old,  $\ln W$ . It satisfies condition (a) because it is measured in dollars (or some other currency that can be exchanged for real goods, such as apples). According to the research findings of Heckman, Stixrud, and Urzua (2006) it also satisfies (b) as higher cognitive and noncognitive skills cause  $\ln W$  to increase. We can model the relationship between the natural logarithm of wages at age 30 and cognitive and noncognitive skills at age 13 as:

$$\ln W_i = \alpha_N \theta_{i,T}^N + \alpha_C \theta_{i,T}^C + \nu_{i,t}. \quad (16)$$

Cunha, Heckman, and Schennach (2006) call equations such as (16), anchoring equations and the functions:

$$g_N(\theta_{i,T}^N) = \alpha_N \theta_{i,T}^N$$

$$g_C(\theta_{i,T}^C) = \alpha_C \theta_{i,T}^C$$

anchoring functions. These functions can be used to transform information from scores into dollar figures. The anchor functions allow us to estimate the technology functions (12) and (13) by measuring skills according to a dollar metric, which is meaningful.

Many different outcomes may serve as anchors, because by definition, any outcome that has a natural metric and is correlated with skills that can serve as anchors. The natural logarithm of wages is only one possibility. Another possibility is the probability of graduating from high school. If we model high school graduation according to a linear probability model, then the same steps above apply. However, the anchoring equations and anchoring functions need not be linear relationships. They can be nonlinear, although we do not develop them here.

It is helpful to establish a simple notation so we can discuss the various estimation procedures applied by Cunha, Heckman, and Schennach (2006) for the type approach we have just explicated. With that in mind, let the observed data for child  $i$  at stage  $t$  be stacked in vector  $Z_{i,t}$  in the following manner:

$$Z_{i,t} = [Z_{i,t,1}^C, Z_{i,t,2}^C, Z_{i,t,1}^N, Z_{i,t,2}^N, Z_{i,t,1}^I, Z_{i,P,1}^C, Z_{i,P,2}^C, Z_{i,P,1}^N, Z_{i,P,2}^N]' .$$

We can also stack in the vector  $\theta_{i,t}$  the latent parents' and children's skills as well as parental investments:

$$\theta_{i,t} = [\theta_{i,t}^C, \theta_{i,t}^N, I_{i,t}, \theta_P^C, \theta_P^N] .$$

Similarly, we organize in the same vector  $\varepsilon_{i,t}$  the noise from observations  $Z_{i,t}$  :

$$\varepsilon_{i,t} = [\varepsilon_{i,t,1}^C, \varepsilon_{i,t,2}^C, \varepsilon_{i,t,1}^N, \varepsilon_{i,t,2}^N, \varepsilon_{i,t,1}^I, \varepsilon_{i,P,1}^C, \varepsilon_{i,P,2}^C, \varepsilon_{i,P,1}^N, \varepsilon_{i,P,2}^N]' .$$

We can then write the system that links observed data  $Z_{i,t}$  to unobserved latent variables  $\theta_{i,t}$  and noise

$\varepsilon_{i,t}$  in the following manner:

$$Z_{i,t} = \alpha_t \theta_{i,t} + \varepsilon_{i,t}. \quad (17)$$

To complete the model, we rewrite the technology equations to allow for modelling errors  $\eta_{i,t}^C$  and  $\eta_{i,t}^N$  in the following fashion:

$$\theta_{i,t+1}^C = f_t^C (\theta_t^C, \theta_t^N, I_t, \theta_P^C, \theta_P^N) + \eta_{i,t}^C \quad (18)$$

$$\theta_{i,t+1}^N = f_t^N (\theta_t^C, \theta_t^N, I_t, \theta_P^C, \theta_P^N) + \eta_{i,t}^N \quad (19)$$

In the literature in statistics, models that combine equations (18), (19), and (17) are called dynamic factor models, or, more generally, state space models. The estimation of dynamic factor models is usually done by a filtering technique, such as the popular Kalman Filter. Filtering is a technique that allows the researcher to explore the recursiveness of the problem to attain considerable improvements in computational speed, which is attained by factorizing the likelihood function. Let  $p(Z_{i,1}, \dots, Z_{i,T})$  denote child's  $i$  contribution to the likelihood. This is a function that has many arguments. Filtering techniques, such as the Kalman Filter, allow the analyst to write:

$$p(Z_{i,1}, \dots, Z_{i,T}) = p(Z_{i,T} | Z_{i,1}, \dots, Z_{i,T-1}) \times p(Z_{i,T-1} | Z_{i,1}, \dots, Z_{i,T-2}) \times \dots \times p(Z_{i,1})$$

where  $p(Z_{i,s} | Z_{i,1}, \dots, Z_{i,s-1})$  is the probability density function of  $Z_{i,s}$  conditional on the observed history up to period  $s$ :  $Z_{i,1}, \dots, Z_{i,s-1}$ . This factorization provides gains in computational speed because it replaces the evaluation of a function of many arguments by a simple multiplication of many functions of one argument.

The Kalman Filter can be applied if the following five conditions are satisfied. First, the relationship between  $Z_{i,t}$ ,  $\theta_{i,t}$ , and  $\varepsilon_{i,t}$  has to be linear and separable as described in (17). Second, the variables  $\theta_{i,t}$  must be independent from the noise variables  $\varepsilon_{i,t}$  and modelling errors  $\eta_{i,t} = (\eta_{i,t}^C, \eta_{i,t}^N)$ . Third, the noise variable  $\varepsilon_{i,t}$  must be independent from the modelling errors  $\eta_{i,t}$ . Fourth, the variable  $\theta_{i,t}$ ,  $\varepsilon_{i,t}$ , and  $\eta_{i,t}$  are required to be jointly normally distributed. Fifth, the technology functions (18) and (19) must be linear so that the the technology for noncognitive skills is described by

$$\theta_{i,t+1}^N = B_{t,N} + \gamma_{t,N,1} \theta_{i,t}^C + \gamma_{t,N,2} \theta_{i,t}^N + \gamma_{t,N,3} I_{i,t} + \gamma_{t,N,4} \theta_{i,t}^C + \gamma_{t,N,5} \theta_{i,t}^C + \eta_{i,t}^N,$$

and the technology for cognitive skills is modelled according to

$$\theta_{i,t+1}^C = B_{t,C} + \gamma_{t,C,1}\theta_{i,t}^C + \gamma_{t,C,2}\theta_{i,t}^N + \gamma_{t,C,3}I_{i,t} + \gamma_{t,C,4}\theta_{i,P}^C + \gamma_{t,C,5}\theta_{i,P}^C + \eta_{i,t}^C.$$

These assumptions are restrictive, and rule out models with general substitution among investments of various types and stages. Nonetheless, with this technology one can still estimate the self-productivity parameters  $\gamma_{t,N,1}$  and  $\gamma_{t,C,2}$  and gauge the importance of self-productivity. These assumption also allow us to test for the presence of cross-productivity in the accumulation of skills, which are given by the parameters  $\gamma_{t,N,2}$  (current cognitive skills increase the stocks of future noncognitive skills) and  $\gamma_{t,C,1}$  (current noncognitive skills increase the stocks of future cognitive skills). The model is sufficiently flexible to understand how parental investments affect the accumulation of cognitive versus noncognitive skills: one just has to compare the estimated values (and standard errors) of  $\gamma_{t,N,3}$  with those for  $\gamma_{t,C,3}$ . One can study how the effect of parental investments varies over time in the accumulation of a given skill, which can be done by comparing  $\gamma_{t,N,3}$  against  $\gamma_{\tau,N,3}$  or  $\gamma_{t,C,3}$  against  $\gamma_{\tau,C,3}$  for  $t \neq \tau$ . One can also use the technology to investigate the impact of parental skills in the accumulation of the skills of children. Furthermore, one can estimate the vector of factor loadings  $\alpha_t$  and obtain optimal weights which are used for the construction of the parental investment variable  $I_t$ , as described above.

The major limitation of using Kalman Filtering is that it does not allow us to estimate the degree of complementarity, which is given by the parameters  $\phi_N$  and  $\phi_C$ , respectively. Given the practice in the literature in the economics of education, as illustrated by Todd and Wolpin (2003, 2005), we start by estimating linear technologies. We will answer all the questions that can be answered in such a framework. In the end, we will present an alternative estimation strategy that will allow us to recover these parameters, but at a much greater computational cost.

## 7.1 Estimation of Normal-Linear Technologies Using the Kalman Filter

Cunha and Heckman (2006) use a sample of the 1053 white males from the Children of the NLSY/79 (CNLSY/79) data set. Starting in 1986, the children of the NLSY/1979 female respondents have been assessed every two years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as their home environment.

Data were collected via direct assessment and maternal report during home visits at every biannual wave. Tables 8A and 8B presents summary statistics of our data.

The measures of quality of a child's home environment that are included in the CNLSY/79 survey are the components of the Home Observation Measurement of the Environment - Short Form (HOME-SF). They are a subset of the measures used to construct the HOME scale designed by Bradley and Caldwell (1980, 1984) to assess the emotional support and cognitive stimulation children receive through their home environment, planned events and family surroundings. These measurements have been used extensively as inputs to explain child characteristics and behaviors (see e.g. Todd and Wolpin, 2005). As discussed in Linver, Brooks-Gunn, and Cabrera (2004), some of these items are not useful because they do not vary much among families (i.e., more than 90% to 95% of all families make the same response). Our empirical study uses measurements on the following parental investments: the number of books available to the child, a dummy variable indicating whether the child has a musical instrument, a dummy variable indicating whether the family receives a daily newspaper, a dummy variable indicating whether the child receives special lessons, a variable indicating how often the child goes to museums, and a variable indicating how often the child goes to the theater. Cunha and Heckman (2006) also report results from some specifications that use family income as a proxy for parental inputs, but none of their empirical conclusions rely on this particular measure.

As measurements of noncognitive skills we use components of the Behavior Problem Index (BPI), created by Peterson and Zill (1986), and designed to measure the frequency, range, and type of childhood behavior problems for children age four and over, although in our empirical analysis we only use children age six to thirteen. The Behavior Problem score is based on responses from the mothers to 28 questions about specific behaviors that children age four and over may have exhibited in the previous three months. Three response categories are used in the questionnaire: often true, sometimes true, and not true. In their empirical analysis they use the following subscores of the behavioral problems index: (1) antisocial, (2) anxious/depressed, (3) headstrong, (4) hyperactive, (5) peer problems. Among other characteristics, a child who scores low on the antisocial subscore is a child who often cheats or tell lies, is cruel or mean to others, and does not feel sorry for misbehaving. A child who displays a low score on the anxious/depressed measurement is a child who experiences sudden changes in mood, feels no one loves him/her, is fearful, or feels worthless or inferior. A child with low scores on the headstrong measurement is tense, nervous, argues

too much, and is disobedient at home, for example. Children will score low on the hyperactivity subscale if they have difficulty concentrating, act without thinking, and are restless or overly active. Finally, a child will be assigned a low score on the peer problem subscore if they have problems getting along with others, are not liked by other children, and are not involved with others.

For measurements of cognitive skills we use the Peabody Individual Achievement Test (PIAT), which is a wide-ranging measure of academic achievement of children aged five and over. It is widely used in developmental research. Todd and Wolpin (2005) use the raw PIAT test score as their measure of cognitive outcomes. The CNLSY/79 includes two subtests from the full PIAT battery: PIAT Mathematics and PIAT Reading Recognition<sup>41</sup>. The PIAT Mathematics measures a child's attainment in mathematics as taught in mainstream education. It consists of 84 multiple-choice items of increasing difficulty. It begins with basic skills such as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The PIAT Reading Recognition subtest measures word recognition and pronunciation ability. Children read a word silently, then say it aloud. The test contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include the ability to match letters, name names, and read single words aloud.

### 7.1.1 Estimates of Time-Invariant Linear Technology Parameters

Using the CNLSY data, we first estimate the simplest version of the model that imposes the restriction that the coefficients on the technology equations do not vary over periods of the child's life cycle. We first report results in the scale of standardized test scores. Below, we show the estimated technology using an anchoring function.<sup>42</sup>

Table 9 shows the estimated parameter values and their standard errors. From this table, we see that: (1) both cognitive and noncognitive skills show strong persistence over time; (2) noncognitive skills affect the accumulation of next period cognitive skills, but cognitive skills do not affect the accumulation of next period noncognitive skills; (3) the estimated parental investment factor affects noncognitive skills somewhat more strongly than cognitive skills, although the differences are not statistically significant; (4) the mother's cognitive ability affects the child's cognitive ability but not noncognitive ability; (5) the

---

<sup>41</sup>We do not use the PIAT Reading Comprehension battery since it is not administered to the children who score low in the PIAT Reading Recognition.

<sup>42</sup>See Cunha and Heckman (2006) for the effect of alternative normalizations on our estimates.

mother's non-cognitive ability affects the child's noncognitive ability, but not the child's cognitive ability. These results are robust to alternative normalizations of the factor loadings on the measurements associated with family inputs that set the scale of the parental investment factor as discussed in Cunha and Heckman (2006).

The strong self-productivity for cognitive skills as reported in our estimates from Table 9 has been shown before by other researchers such as Todd and Wolpin (2005). We find that the same pattern holds for the accumulation of noncognitive skills as well. We are unaware of studies that model the evolution of cognitive and non-cognitive jointly, so the finding of cross-productivity effects from noncognitive skills to cognitive skills is to the best of our knowledge a new finding.

We are also unaware of studies which compare the impact of parental investments in the accumulation of noncognitive versus cognitive skills, although the studies in biology do point to a similar findings and the explanation is that noncognitive skills are associated with the prefrontal cortex whose development process is not finished until the person is already beyond 20 years-old.

The fact that mother's cognitive skills increase children's cognitive skills and mother's noncognitive skills increase children's noncognitive skills have both been shown in a different context by Duncan, Kalil, Mayer, Tepper, and Payne (2005). They show that mother's cognitive test scores can predict children's cognitive test scores. Furthermore, mother's behaviors (such as smoking) which correlate with mother's cognitive skills, is a good predictor of the same behavior for the children.

The dynamic factors are estimated to be statistically dependent. Table 10 shows the evolution of the correlation patterns across the dynamic factors. Early in the life cycle, the correlation between cognitive and noncognitive skills is strong. The correlation is 0.21 as early as ages 6 and 7, and it grows to around 0.29 at ages 12 and 13. There is also strong contemporaneous correlation among noncognitive skill and the home investment. The correlation starts off at 0.29 at ages 6 and 7 and grows to 0.45 by ages 12 and 13. The same pattern is true for the correlation between cognitive skills and home investments. In fact, the correlation between these two variables actually doubles from 0.26 at ages 6 and 7 to 0.35 at ages 12 and 13.

### 7.1.2 Anchoring our estimates of the factor scale using adult outcomes

We now report estimates that use the probability of graduating from high-school data for persons age 19 and above to anchor the output of the production function in an interpretable metric. For the CNLSY data, using high-school graduation is more indicated as observation of labor earnings data is only available for children who come from very disadvantaged families<sup>43</sup>. Our fitted probability model is a linear probability model on year of birth of the child, and the final level of the factors  $\theta_T^C$  and  $\theta_T^N$ . The coefficient on cognitive skills in the log earnings equations is estimated to be 0.14 (standard error is 0.054). For noncognitive skills, we estimate a loading of 0.052 (with a standard error of 0.0109). These estimates are consistent with estimates reported in Heckman, Stixrud, and Urzua (2006).

Table 11, which transforms the estimates in Table 9 into a high-school graduation metric, shows that some of our conclusions are altered when we anchor outcomes in an adult outcome rather than just in a test score. The most important effect is that in the metric of high school graduation, the impact of parental investments is greater on the accumulation of cognitive than noncognitive skills.

### 7.1.3 Evidence of Sensitive Periods of Investment in Skills

We provide evidence supporting the existence of sensitive periods in the accumulation of cognitive and noncognitive skills. We can identify whether there are sensitive periods in the development of skills provided that we normalize our investment factor on an input that is used at all stages of the life cycle. Using several alternative measures including trips to the theater, the number of books, as well as family income as a “proxy,” we obtain the same qualitative ordering in terms of critical and sensitive periods. All of our estimated models include an equation for the probability of graduating from high-school of the child based on the period  $T$  value of the factors but “output” is reported in test score units.

Using a likelihood ratio test, we test and reject the hypothesis that the parameters describing the technologies are invariant over stages of the lifecycle. Specifically, we use a likelihood ratio test. Under the restricted model, we estimate 277 parameters and the value of the log likelihood at the maximum is -53877. Under the unrestricted model, we estimate 305 parameters and the log likelihood attains the maximum

---

<sup>43</sup>This problem arises because the CNLSY is not a representative sample of U.S. children population, since it is surveys the children of the NLSY/1979 respondents. The NLSY/1979 respondents are a representative sample of the U.S. population born between 1957 and 1964.

value of -53800. The statistic  $\lambda = -2(\ln L_R - \ln L_U)$  is distributed as chi-square with 28 ( $= 305 - 277$ ) degrees of freedom. We find that  $\lambda$  is 155, significantly above the critical value of 41.337 at a 5% significance level.

Our estimates are reported in Table 12. When we allow the coefficients of the technology to vary over age we find evidence of sensitive periods for both cognitive and noncognitive skills. A sensitive period for parental investments in cognitive skills occurs at an earlier age than the sensitive period for parental investments in noncognitive skills. The coefficient on investments in the technology for cognitive skills for the transition from period one to period two (ages 6 and 7 to ages 8 and 9) is around 0.072 (with a standard error of 0.0152). For the transition from period two to period three (ages 8 and 9 to 10 and 11) this same coefficient decreases rather sharply to 0.0178 (with a standard error of 0.0061). For the final transition (ages 10 and 11 to ages 12 and 13), this coefficient is about 0.0160, with a standard error of 0.0073. The difference between the early coefficient and the later two is statistically significant. This finding is consistent with periods 1 and 2 being sensitive periods for cognitive skills.<sup>44</sup>

For noncognitive skills in period one, the coefficient on investments is only 0.0204, with a standard error of 0.0101. Then, it increases to 0.0593 in period two (with standard error of 0.0206). At the last transition, this coefficient is 0.1038 with a standard error of 0.0213. This evidence suggests that the sensitive periods for the development of noncognitive skills tend to take place at later ages in comparison to sensitive periods for cognitive skills.<sup>45</sup>

#### 7.1.4 Estimating the Components of the Home Investment Dynamic Factor

In Table 13 we show how our method constructs an implicit home score by estimating factor loadings on the inputs used to form the conventional home score. We use the estimates generating the parameters reported in Table 11. Thus we normalize the scale of the investment factor by “trips to the theater”. The CNLSY/1979 reports an aggregate HOME score by adding these variables and assigning each one of them the same weight. For expositional purposes we call these ad-hoc weights. The advantage of working with (dynamic) factor models is that the relative weights on the components of the home score are estimated

---

<sup>44</sup>For the coefficients on cognitive skills, the lower bound for the  $t$  statistic for the hypothesis  $\gamma_{I,2}^C = \gamma_{I,1}^C$  is 2.73. For the hypothesis  $\gamma_{I,2}^C = \gamma_{I,3}^C$  it is 3.43.

<sup>45</sup>For the coefficients of investments on noncognitive skills, the lower bound for the  $t$  statistic for the hypothesis  $\gamma_{I,2}^N = \gamma_{I,1}^N$  is 2.16 and for the hypothesis  $\gamma_{I,2}^N = \gamma_{I,3}^N$  it is 2.34.

rather than imposed, as we derived above. We can also test how informative each component is in forming the parental investment variable. This feature makes the approach applied here appealing as one can test how well certain variables proxy the unobservable variable of interest.

For example, consider the number of books available to the child. In general, we should expect this variable to be correlated with parental inputs because parents who invest more in the development of their children will tend to spend more resources on books. But the number of books is unlikely to be a perfect indicator of total parental input. Our method allows for imperfect proxies. Under our method, the number of books child  $i$  has at age  $t$  (say,  $Z_{i,t,1}^I$ ) is modelled as

$$Z_{i,t,1}^I = \alpha_{t,1}^I I_{i,t} + \varepsilon_{i,t,1}^I.$$

Under the assumption of independence between  $I_{i,t}$  and  $\varepsilon_{i,t,1}^I$ , it follows that

$$\text{Var} (Z_{i,t,1}^I) = (\alpha_{t,1}^I)^2 \text{Var} (I_{i,t}) + \text{Var} (\varepsilon_{i,t,1}^I).$$

We can decompose the total unobserved variance in two terms: one that is due to the parental input,  $s_{i,t,1}^I$ , and another that is due to noise,  $\tilde{s}_{i,t,1}^I$ . The relative importance of the two measures can be computed as:

$$s_{i,t,1}^I = \frac{(\alpha_{t,1}^I)^2 \text{Var} (I_t)}{(\alpha_{t,1}^I)^2 \text{Var} (I_t) + \text{Var} (\varepsilon_{i,t,1}^I)},$$

and

$$\tilde{s}_{i,t,1}^I = \frac{\text{Var} (\varepsilon_{i,t,1}^I)}{(\alpha_{t,1}^I)^2 \text{Var} (I_t) + \text{Var} (\varepsilon_{i,t,1}^I)}.$$

Table 13 reports that  $\tilde{s}_{i,t,1}^I = 0.1359$  (corresponding to 6 and 7), while  $s_{i,t,1}^I = 0.8641$ . So, most of the unobservable variance in “the number of books a child has” is actually not informative on the parental input unobserved variable  $I_t$ . We report the same measures for the other input variables in Table 13. Over stages of the life cycle, all of the input measures tend to become more error laden as a proxy for  $I_t$ . Using this procedure, it is possible to identify at each age of the child, what are the inputs that are most important in producing effective investments. For example, books are better informative about parental investments early on than later on. One explanation is as children spend more and more years at school

the number of books they acquire converge to a distribution with little variation, so this measure becomes a poor measure in terms of differences in parental investments.

Another way to see this is by the constructed weights. Table 13 displays the estimated weights  $w_{k,t}$  for each measurement  $k$  at each period  $t$ . Note that the weights are not stable over stages of the life cycle. Our estimates show that the number of books receives high weight early on (ages 6/7 and 8/9), but the weight declines considerably in the later periods (ages 10/11 and 12/13). The variable that indicates whether the child receives special lessons, on the other hand, exhibits the opposite behavior. It starts small in early ages, but it becomes more important at later ages. It is interesting to remark that variables that describe how often children attend theater or visit museums, although informative about the home investments, receive lower weights in our method than the weights that weight all items equally strongly.

### 7.1.5 Estimation of Non-Linear Technologies

Because of the nonlinearity of our general model we cannot use Kalman filtering. We use particle filtering methods to obtain  $p(Z_{i,t}|Z_{i,1}, \dots, Z_{i,t-1})$  for  $t = 2, \dots, T$  (see Doucet, de Freitas, and Gordon, 2001; Hammersley and Morton, 1954).

Note that:

$$\begin{aligned} p(Z_{i,t}|Z_{i,1}, \dots, Z_{i,t-1}) &= \int p(Z_{i,t}, \theta_t|Z_{i,1}, \dots, Z_{i,t-1}) d\theta_t = \\ &= \int p(Z_{i,t}|\theta_t, Z_{i,1}, \dots, Z_{i,t-1}) p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1}) d\theta_t \\ &= \int p(Z_{i,t}|\theta_t) p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1}) d\theta_t. \end{aligned}$$

Thus,

$$p(Z_{i,t}|Z_{i,1}, \dots, Z_{i,t-1}) = \prod_{t=1}^T \int p(Z_{i,t}|\theta_t) p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1}) d\theta_t. \quad (20)$$

From our assumption about measurement errors, we know that  $p(Z_{i,t}|\theta_t) = p(\varepsilon_{i,t})$ . The problem is to construct  $p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1})$ . Nonlinear filters are algorithms that, given  $p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1})$ , allow one to compute  $p(\theta_{t+1}|Z_{i,1}, \dots, Z_{i,t})$ . Similar to the Kalman filter, nonlinear filtering breaks this task into two steps: update and prediction. The update step produces  $p(\theta_t|Z_{i,1}, \dots, Z_{i,t})$  given  $p(\theta_t|Z_{i,1}, \dots, Z_{i,t-1})$ .

To perform this update step apply Bayes' rule:

$$p(\theta_t | Z_{i,1}, \dots, Z_{i,t}) = \frac{p(Z_{i,t} | \theta_t) p(\theta_t | Z_{i,1}, \dots, Z_{i,t-1})}{p(Z_{i,t} | Z_{i,1}, \dots, Z_{i,t-1})}, \quad (21)$$

where the normalizing constant is  $p(Z_{i,t} | Z_{i,1}, \dots, Z_{i,t-1})$  which depends on  $p(z_t | \theta_t) = p(\varepsilon_t)$  as defined by the measurement equation.

The prediction step generates  $p(\theta_{t+1} | Z_{i,1}, \dots, Z_{i,t})$  given  $p(\theta_t | Z_{i,1}, \dots, Z_{i,t})$ , using the technology functions (18) and (19) to obtain the prediction density of  $\theta_t$  using the Chapman-Kolmogorov equation:

$$p(\theta_{t+1} | Z_{i,1}, \dots, Z_{i,t}) = \int p(\theta_{t+1} | \theta_t) p(\theta_t | Z_{i,1}, \dots, Z_{i,t}) d\theta_t. \quad (22)$$

By combining update and prediction steps, one can calculate  $p(\theta_{t+1} | Z_{i,1}, \dots, Z_{i,t})$  given  $p(\theta_t | Z_{i,1}, \dots, Z_{i,t-1})$  and we can write the likelihood recursively.<sup>46</sup>

Cunha, Heckman, and Schennach (2006) use a *CES* technology applied to the proxy data on investment and latent skills to estimate a time-invariant version of the technology of skill formation. The normalizations on the measurements we use are presented in our online tables. The specific form of the *CES* technology they use is

$$\theta_{t+1}^{k+1} = B_k \left[ \gamma_{k,1} (\theta_t^N)^{\phi_k} + \gamma_{k,2} (\theta_t^C)^{\phi_k} + \gamma_{k,3} (I_t)^{\phi_k} + \gamma_{k,4} (\theta_M^C)^{\phi_k} + \gamma_{k,5} (\theta_M^N)^{\phi_k} \right]^{1/\phi_k} \exp(\eta_t^k), \quad (23)$$

where  $\sum_{l=1}^5 \gamma_{k,l} = 1$ . The factors are required to be nonnegative to define the technology.

We report both anchored results and unanchored results, using the nonlinear version of anchoring described in detail in Cunha, Heckman, and Schennach (2006). The anchored results allow us to compare the productivity of investments and stocks of different skills at different stages of the life cycle on the anchored outcome. We first report results in the scale of standardized test scores. We discuss estimates in the scale of the probability of graduating from high school below.

Table 14 shows the estimated parameter values and their standard errors in the unanchored system. From this table, we see that: (1) both cognitive and noncognitive skills show strong persistence over time; (2) noncognitive skills affect the accumulation of the next period's cognitive skills and cognitive skills

---

<sup>46</sup>Further details on our implementation of particle filtering are presented in Cunha, Heckman, and Schennach (2006).

affect the accumulation of the next period's noncognitive skills; (3) the estimated parental investment factor affects noncognitive skills slightly more strongly than cognitive skills, although the differences are not statistically significant; (4) the mother's ability affects both the child's cognitive and noncognitive ability; (5) the mother's noncognitive skills also affect test outcomes.

These results differ from those given by the linear technology previously reported, where we impose the condition  $\phi_C = \phi_N = 1$  in estimating the model. In the linear model we found no role for the mother's ability on child noncognitive skill, whereas in the nonlinear model we find a stronger role. We find that cognitive skills affect the accumulation of next period noncognitive skills. Allowing for general forms of substitution affects the estimates.

The elasticities of substitution between investments and stocks of skills are both below 1, with noncognitive investments technologically more substitutable across periods than cognitive investments. This finding is consistent with the evidence on plasticity of noncognitive skills and the lesser plasticity of cognitive skills discussed in Part II.

To circumvent the problem that test score units are intrinsically arbitrary, Cunha, Heckman, and Schennach (2006) anchor outcomes in terms of their effect on high school graduation. Table 15 reports anchored estimates in the probit of high school graduation. Compared to the unanchored case, anchoring increases the estimated elasticity of substitution for both estimated skills, especially for noncognitive skills. Both estimates are still below 1 ( $\phi_C \cong -0.25$ ,  $\phi_N \cong -0.12$ ).<sup>47</sup> The qualitative conclusions of Table 14 survive. In the anchored case, we can meaningfully compare the effects of parental investments on childhood outcomes. It is interesting to see that parental investments have similar impacts on cognitive and noncognitive skills once we anchor on the probability of graduating from high school.

---

<sup>47</sup>When we use the linear probability model, the estimated elasticities of substitution are slightly larger. See our website <http://jenni.uchicago.edu/elastic-sub/>.

## Part V

# Simulating the Estimated Model: Validation and Lessons for the Design of Policies

## 8 Model Validation - Out of Sample Predictions

At this point we have assembled a large array of evidence on the formation of cognitive and noncognitive skills. On one hand, sensitive periods for cognitive skills take place early in childhood. On the other hand, noncognitive skills are subject to sensitive periods later in life. From the analysis of the nonlinear models, we confirmed the evidence that it is easier to substitute investments intertemporally for noncognitive skills than cognitive skills. Note that the linear model is nested in the nonlinear model. The fact that the data is indicating for the presence of nonlinearities is already evidence against the linear models. However, we can go further and test the validity of the linear and nonlinear models using out of sample predictions.

There are not many data sets that contain the rich information used in the estimation of the technologies. However, one can use data from the Perry Preschool experiment to check which model better predicts the evolution of cognitive skills: the linear model or the nonlinear model?

The first problem is that we don't have information on mother's cognitive and noncognitive skills from Perry Preschool. This is a limitation, because our estimates show that mothers' skills affect the accumulation of skills of the children. In the absence of any information, we assume that the mothers from Perry are at the bottom quartile of the distribution of mother's skills. This is a reasonable assumption for the Perry sample because this was a program targeted to very disadvantaged households.

The second problem is that we don't observe the same variables we used for the construction of our parental investment variables in the Perry data. However, we can obtain an estimate of investments for control and treatment group under some assumptions. For example, we know that both control and treatment groups are around the first decile in the distribution of cognitive skills at age of entry.<sup>48</sup> We also

---

<sup>48</sup>At age of entry, the average score on the Stanford Binet cognitive test was around 80. The population mean is 100 with a standard deviation of 16. If we assume that the Stanford Binet scores follow a normal distribution, one obtains that 80 is around the first decile in the distribution of Stanford Binet.

know that the treatment group at the program exit age is around the fourth decile in the distribution of cognitive skills. Because we have no information on noncognitive scores at entry and exit ages, we make the assumption that at age of entry the children are at the first decile in the distribution of noncognitive skills. We make no assumptions about the location of the children in the distribution of noncognitive skills at the age of exit from the program.

Assume that children's initial conditions are the same as at the age of entry into the program. Both treatment and control groups are born and enter at the lowest decile of cognitive and noncognitive skills. We can estimate the level of investment received after birth by the control group by finding the level of investment, say  $I_{cont}$ , such that children born with first decile cognitive and noncognitive skills are also in the first decile of cognitive and noncognitive skills at the age of entry into the program. Given our estimated parameter values, we find that  $I_{cont}$  is around the bottom quartile in the distribution of investments, for both linear and nonlinear models. Since Perry participants form a fairly disadvantaged group, this number seems like a reasonable estimate.

We can then ask for the treatment group what the level of investment is during the treatment years, say  $I_{treat}$ , by matching the scores on cognitive skills tests at age of exit from the program. At age five, the average score in Stanford Binet was around 95 points, which is roughly at the fourth decile in the distribution of cognitive skills. For the nonlinear model, we find that  $I_{treat}$  is around the seventh decile, while for the linear model we estimate that  $I_{treat}$  is at the top decile. This is evidence against the linear model. Although Perry was indeed an enriched early environment program, it was not full time, for example, and is unlikely to have had such a substantial increase in investments.

We then simulate the time series of average test scores for control and treatment groups of a Perry Preschool program using both the linear and nonlinear models and we compare against actual Perry data. In Figure 11, we plot the time series of average cognitive scores for the actual treatment group (solid curve) against the simulated Perry from the linear model (dashed curve) and the simulated Perry from the nonlinear model (dashed-dotted curve). As explained above, the level of investments  $I_{cont}$  and  $I_{treat}$  are such that both linear and nonlinear models predict exactly the scores at age 3 (the age of entry) and at age 5 (the age of exit). There is one feature similar between the linear and nonlinear models: they both exhibit a sluggish response. While in the actual Perry data, the score on the Stanford Binet is 97 points, the linear model predicts a score of only 87.5 points and the nonlinear model predicts a score of 87.4 points.

This discrepancy may be the result of sensitive periods in the accumulation of cognitive skills. Using a technology based on time-invariant parameters, we would be underestimating the impact of investments in cognitive skills at early ages.

However, the linear model predicts a much faster decay in test scores for the treatment group than the actual Perry treatment group experienced and the simulated treatment from the nonlinear model. In fact, at the end of our sample (age 10), the average score for the actual Perry treatment and the simulated one from the nonlinear model almost coincide: 85.3 for the actual Perry treatment against 85.5 for the simulated nonlinear model. The linear model predicts an average score for the treatment group around 80 points.

The performance of the linear model is even worse for the control group. Figure 12 plots average Stanford Binet scores for the Perry control group against the simulated linear (dashed curve) and nonlinear (dashed-dotted curve) models. In the Perry data, the scores in Stanford Binet increase from roughly 80 points at age 3 to almost 88 points at age 7 (the peak for the control in Figure 12). The prediction of the linear model cannot even fit the trend: it predicts a decrease to roughly 70 points in Stanford Binet at age 7, which is a score at borderline mental retardation. The nonlinear model does not capture the rapid increase in test scores actually observed, but on the other hand, it does not predict a decrease in scores either. According to the nonlinear model, at age 7 the control group would score a little more than 81 points in Stanford Binet. At age 10, the nonlinear model predicts a score of 82.7 points, while in the actual data the average score is 84. The linear model predicts a score of only 66 points. This is again evidence against the linear models, which assume that investments are perfectly substitutable over ages.

## **9 How Early Environments Promote Education and Reduce Crime**

We simulate the model to show how early investment, followed up by complementary later investment, reduces crime and promotes educational attainment. To understand the significance of these results it will be useful to present a context on trends in education and the consequences of improving education.

Table 16, taken from Ellwood (2001), highlights a major problem facing the American labor market in

the next two decades. The first column of the table presents the distribution of the American workforce among age and race-ethnicity categories in 1980. The second column shows the growth in the categories from 1980 to 2000 and the third column shows the labor force as of 2000. The fourth column shows the projected growth in the labor force in the next twenty years by category. Except for the numbers for immigrants, these are reliable projections because there is little emigration and the groups being projected are already alive. The immigration projections come from a carefully executed U.S. Census study. The labor force is aging and young replacements for old workers are increasingly in short supply compared to the 1980s. The aging of the American workforce raises serious problems for the future of American productivity growth.

The workforce of prime-age workers, fueled by the entry of Baby Boomers, propelled U.S. economic growth in the period 1980–2000. However, we cannot count on this source of growth in the next twenty years. Indeed, the largest components of growth in the workforce will come from older workers as the Baby Boom cohort ages. A major source of vitality in the U.S. workforce will be lost. Future workforce growth will come from older workers and from demographic groups in which, for a variety of reasons, dysfunctional and disadvantaged families are more prevalent.

On top of these trends in the number of workers by age, there is stagnation in educational attendance rates. Figure 13 shows the distribution of educational attainment among 30-year-olds by year. College-going rates have stalled out for cohorts of Americans born after 1950. This is not a consequence of immigration of unskilled workers. It is a phenomenon found among native-born Americans. Currently, 17% of all new high school credentials issued are to GEDs.<sup>49</sup> Heckman (2004) documents that the high school dropout rate has increased over time if one counts GEDs as dropouts. This is appropriate because GEDs earn the same wages as dropouts.

The growth in the quality of the workforce, which was a mainstay of economic growth until recently, has diminished. Assuming that these trends continue, the U.S. economy will add many fewer educated persons to the workforce in the next two decades than it did in the past two decades (see Table 17). Jorgenson, Ho, and Stiroh (2003) estimate that the average annual rate of growth of college labor supply was 4.5% in 1977, but fell to 1.75% in 1990–2000. These trends are predicted to continue, or possibly worsen.

The slowdown in labor force quality growth has already hurt American productivity growth. Delong,

---

<sup>49</sup>The GED is an exam-certified, alternative high school degree.

Katz, and Goldin (2003) estimate that increases in educational attainment boosted the effective quality of the workforce by 0.5% a year over the period 1915–2000, and thus contributed an average of 0.35 percentage points per year to economic growth over the period.<sup>50</sup> The slower growth in educational attainment of the workforce substantially reduced productivity growth compared to that experienced in the 1915–1980 era. Based on current trends, these authors project that the annual rate of productivity growth attributable to education—0.35 from 1980 through 2000—will decline by half or more (to between 0.06 and 0.17 percent) in the next two decades. This will reduce the productivity growth of labor by a substantial 0.18–0.29 percentage points per year and will be a drag on real wage growth and on fiscal revenues.

## 9.1 Literacy and Numeracy

The skills of the U.S. labor force are poor. The U.S. has a thick lower tail of essentially illiterate and innumerate persons, who are a drag on productivity and a source of social and economic problems. We use data from the International Adult Literacy Survey (IALS) to examine literacy and numeracy of adults of working age (16–65 years).<sup>51</sup> Document literacy is defined as the ability to locate and use information from timetables, graphs, charts and forms. We present data on document literacy in Figure 14. Tests for prose literacy and quantitative literacy produce the same pattern. See Figures 15 and 16.<sup>52</sup>

Level 1 performance is essentially functional illiteracy or innumeracy: it represents the inability to determine the correct amount of medicine from information on the package. People who perform at Level 1 can make limited use of texts that are simple and uncomplicated. They are only able to locate information in text or data as long as there is no distracting information around the correct answer. On the quantitative scale they can only carry out relatively straightforward operations such as simple addition. Roughly 20% of U.S. workers fall into this category on each test: a much higher fraction than in some of the leading

---

<sup>50</sup>The share of labor is 0.7 so  $0.7 \times 0.5 = 0.35$  is the contribution of workforce quality to economic growth.

<sup>51</sup>The International Adult Literacy Survey (IALS) was conducted by 13 countries to collect information on adult literacy. In this survey, large samples of adults (ranging from 1,500 to 6,000 per country) were given the same broad test of their literacy skills between 1994 and 1996. The participating countries are Australia, Belgium (Flanders), Canada, Germany, Great Britain, Ireland, Netherlands, Northern Ireland, New Zealand, Poland, Sweden, Switzerland and the United States. More information on the IALS is available in documents located at <http://www.nald.ca/nls/ials/introduc.htm> and International Adult Literacy Survey (2002).

<sup>52</sup>Prose literacy is defined as the knowledge and skills required to understand and use information from texts such as newspaper articles and fictional passages. Quantitative literacy is defined as the ability to perform arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as calculating savings from an advertisement or the interest earned on an investment.

European countries. This is a major drag on U.S. competitiveness<sup>53</sup> and a source of social problems.

## 9.2 Crime

Crime is a major burden for American society. Anderson (1999) estimates that the net cost of crime (after factoring out transfers) is over \$1.3 trillion per year in 2004 dollars. The *per capita* cost is \$4,818 per person, in the same dollars. We break down this total in Table 18. This estimate includes crime-induced production (production of personal protection devices, trafficking of drugs and operation of correctional facilities) which costs \$464 billion per year, opportunity costs (production foregone by incarcerated offenders, valued at their estimated wage, time spent locking and installing locks, and so forth) of \$152 billion per year, the value of risks to life and health (pain, suffering and mental distress associated with health losses). This includes time lost from work by victims as well as value of life lost to murders. This component is \$672 billion and is the most controversial item on the list. Yet even ignoring any transfer component, or any risks to life and health, the cost of crime is over 600 billion dollars per year. Although this kind of calculation is necessarily imprecise and there is disagreement over the exact costs, there is widespread agreement that the costs of crime are substantial.

Even though crime rates have recently declined somewhat, their levels remain high. The adult correctional populations (in prison or local jail, on probation or on parole) continue to grow despite the drop in measured crime rates. The size of the population under correctional supervision has continued to grow for all groups, as has the percentage of each group under supervision. Nine percent of blacks were under supervision of the criminal justice system in some form in 1997, although recently this adverse trend has slowed. Incarceration rates have risen steadily since 1980 and only slowed in the late 1990s. The inmate population has risen steadily until recently. Expenditures on prisons, police and the judicial system continue to grow despite the drop in measured crime rates.

These statistics do not convey the full scope of the problem. According to the Bureau of Justice (2004), as of the end of 2001, there were an estimated 5.6 million adults who had ever served time in State or Federal prison: 4.3 million former prisoners and 1.3 million adults in prison. Nearly a third of former prisoners were still under correctional supervision, including 731,000 on parole, 437,000 on probation, and

---

<sup>53</sup>These cross-country differences are not driven by illiterate immigrants. While immigrants perform worse on the three tests relative to natives, including immigrants in the analysis only raises the proportion of US females in Level 1 significantly for prose, quantitative and document literacy. The difference is not significant for any other group or level.

166,000 in local jails. In 2001, an estimated 2.7% of adults in the U.S. had served time in prison, up from 1.8% in 1991 and 1.3% in 1974. The prevalence of imprisonment in 2001 was higher for Black males (16.6%) and Hispanic males (7.7%) than for White males (2.6%). It was also higher for Black females (1.7%) and Hispanic females (0.7%) than White females (0.3%). Nearly two-thirds of the 3.8 million increase in the number of adults ever incarcerated between 1974 and 2001 occurred as a result of an increase in first incarceration rates; one-third occurred as a result of an increase in the number of residents age 18 and older. If recent incarceration rates remain unchanged, it is estimated that one of every 15 persons (6.6%) will serve time in a prison during his or her lifetime.

The lifetime chances of a person going to prison are higher for men (11.3%) than for women (1.8%), and for Blacks (18.6%) and Hispanics (10%) than for Whites (3.4%). Based on current rates of first incarceration, an estimated 32% of Black males will enter state or federal prison during their lifetime, compared to 17% of Hispanic males and 5.9% of White males. Currently, 30% of Black males without a high school degree are in prison (Western, 2006).

What can we do about this problem? One of the best-established empirical regularities in economics is that education reduces crime. Figure 17, from Lochner and Moretti (2004), displays this relationship, reported separately for blacks and whites. Completing high school is a major crime prevention strategy. Poorly educated persons are much more likely to commit crimes than are better educated persons. Other risk factors promoting crime include poor family backgrounds, which also promote dropping out. Poorly educated teenage mothers in low-income families are much more likely to produce children who participate in crime.

Lochner and Moretti (2004) present convincing non-experimental evidence that increasing educational attainment levels reduces crime and that the inverse relationship between crime and education in Figure 17 is not a correlational artifact arising from unobserved variables that are common to both crime and education. Using Census data, they show that 1 more year of schooling reduces the probability of incarceration by 0.37 percentage points for blacks, and 0.1 for whites.<sup>54</sup> To put this evidence in perspective, 23% of the black-white difference in average incarceration rates can be explained by the differences in education between these groups. Using the FBI's Uniform Crime Reports, they find that the greatest impacts of

---

<sup>54</sup>The extra year of school is assumed to take place during high school years. The effect of an extra year of kindergarten or college is likely to be rather different.

education are associated with reducing arrests for murder, assault, and motor vehicle theft.

Lochner and Moretti also calculate the social savings from crime reduction associated with completing secondary education. They show that a 1% increase in the high school graduation rate would yield \$1.8 billion dollars in social benefits in 2004 dollars. This increase would reduce the number of crimes by more than 94,000 in each year (see Table 19). The social benefits include reduced losses in productivity and wages, lower medical costs, and smaller quality-of-life reductions stemming from crime.<sup>55</sup> They also include reductions in costs of incarceration.<sup>56</sup> An increase in *male* high school graduation rates of this magnitude yields a net social benefit of about \$1,638 – 2,967 per additional graduate (in \$2004).

High school graduation confers an extra benefit of 14-26% beyond private returns captured by the high school graduate wages that are pocketed by graduates. This is an important externality that suggests overall under-investment in the population of disadvantaged children at risk for committing crime. Since completing high school raises a student's wages by about \$10,372 per year (in \$2004), and the direct cost of completing one year of secondary school is approximately \$8,000 per student in 1997 (in \$2004), expenditure on schooling is cost-effective. Looking only at the savings from reduced crime, the return is \$1,638 – \$2,967 per year, so that expenditure is cost effective even if we ignore the direct benefits in earnings and even if we assume that the benefits decline as the youths grow older.

Moreover, comparing the effect of educational expenditure with the effect of hiring an additional police officer suggests that promoting education may be a better strategy. Using a somewhat different framework, Levitt (1997) reports that an additional sworn police officer in a large US city would reduce annual costs from crime by about \$200,000 dollars at a public cost of \$80,000 per year. These are recurrent annual costs.

Lochner and Moretti (2004) estimate that in steady state it would cost \$15,000 per year in terms of direct costs to produce enough high school graduates to reduce crime by the same amount. This cost ignores foregone earnings in high school but it also ignores all of the large benefits from high school graduation documented in Heckman, Lochner, and Todd (2004). Educational policy is far more effective per dollar

---

<sup>55</sup>Lochner and Moretti use estimates of victim costs and property losses taken from Miller, Cohen, and Wiersema (1996), which are based on jury awards in civil suits. Some costs cannot be quantified accurately or are unobservable. These include costs of precautionary behavior, private security expenditures, some law enforcement and judicial costs (*i.e.*, costs that are not related to dealing with particular crimes) and the cost of drug offenses. Some crimes are also omitted from the analysis.

<sup>56</sup>Incarceration cost per crime are equal to the incarceration cost per inmate multiplied by incarceration rate for that crime (approximately \$17,000).

spent than expenditure on police.<sup>57,58</sup>

## 10 Lessons for the Design of Policies

We use our estimated nonlinear model of skill formation to show the importance of self-productivity and complementarity for designing policies to promote education and to reduce crime. Using the CNLSY data that are used to fit the nonlinear model, we estimate the impact of interventions at various ages on education and crime.

Let  $\theta_T^N$  and  $\theta_T^C$  denote the stocks of noncognitive and cognitive skills during the adolescent years ( $T = 12, 13$ ). The probability that a person graduates from high school, goes to college, commits a crime, and the like, depends on  $\theta_T^N$ ,  $\theta_T^C$ , and the background variables. Our estimates of the effects of  $\theta_T^N$  and  $\theta_T^C$  on these outcome measures are roughly consistent with those reported in Heckman, Stixrud, and Urzua (2006). Both cognitive and noncognitive skills play important roles in determining these outcomes.

Self-productivity of skills is an important phenomenon. This is illustrated by comparing two different policies. The first policy is a Perry Preschool-like policy. It provides investments at early ages in a way that moves children from the first decile of cognitive skills at entry age to the fourth decile of skills at the age of exit from the program. This is the range of skill enhancement achieved in Perry. Using this information, we find that Perry moved investment in skill from the bottom decile to around the 7th decile of the investment distribution. We also consider a second policy that postpones remediation until adolescence. It compensates early shortfalls by investing larger amounts in adolescence to get roughly the same high school graduation rates observed in Perry.<sup>59</sup> College tuition programs, adolescent literacy programs and mentoring programs are examples. We consider a program in which investments during adolescence are raised to the top decile of the investment distribution. The present value of the costs of the investments in

---

<sup>57</sup>It is important to note that this is a steady state calculation. The payoff to pre-K interventions shows up 10-15 years later, whereas the effects of increasing police on crime are more immediately realized. The discounted returns from the two policies are less different, but a 5:1 gap can tolerate a lot of discounting and still survive.

<sup>58</sup>Lochner and Moretti (2004) actually present a comparison of flow costs (80,000 per year on a police officer) with a one time stock cost (\$600,000 to educate 100 new high school students at a cost of \$6,000 per year assuming that dropouts get 11 years of school. Cameron and Heckman (2001) estimate 10.6 years. Assuming a 40 year working life (including criminal career life) the annual replacement flow cost is \$15,000 a year ( $\$6,000 \times 2.5$ ). Even cutting the career life in half produces a flow cost that is less than hiring a policeman. Spending \$9,000 per year (to account for the 1.5 year gap between high school dropouts and graduates) still makes education cost effective.

<sup>59</sup>When computing high-school graduation rates in Perry we do not consider as high-school graduates individuals who have a GED certification. See Cameron and Heckman (1993) and Heckman and LaFontaine (2006).

the adolescent remediation program is more than 35% larger than in the Perry Preschool program. Late remediation is costly.

We focus our analysis on children from disadvantaged backgrounds because they benefit most from such policies. Disadvantaged children are at risk of being permanently poor and uneducated, and of participating in crime. In our simulations, disadvantaged children come from a background where mothers are in the first decile in the distribution of skills. If no intervention occurs, the children receive investments equivalent to the first decile of the distribution of investments.

The first column in Table 20 reports high-school graduation, college enrollment, conviction, probation, and use of welfare if no intervention is made. Our model predicts a 41% high school graduation rate for this group, compared to 41.4% found in Perry. Only 4.5% of the control group will ever enroll in college. Around 22% of them will be convicted for a crime or be on probation at some point in their adult lives. About 18% will make use of welfare programs in their adult years.

The second column in Table 20 reports the performance of early intervention policies in increasing the welfare of these disadvantaged children. Consistent with the evidence discussed in Part II, this policy increases high school graduation and college enrollment rates to more than 65% and 12%, respectively. It reduces participation in crime. The probability of ever being convicted for a crime or ever being on probation is reduced by around 5.6 and 6.6 percentage points, respectively. It makes the children more productive when they are adults. It cuts in half the probability of collecting welfare benefits in the early adult lives of children.

The third column in Table 20 displays the performance of a 35% more costly policy that produces comparable educational outcomes for those obtained in the Perry-like intervention. Adolescent interventions can work, but they are more costly than early interventions. The greater cost associated with later remediation arises from lost gains in self-productivity from early investments that are a key feature of our model.

The most important difference between the estimates from the linear and the nonlinear models is the role of complementarity. In the estimated linear model, the productivity of investments does not depend on the level of skills. This is not true in the nonlinear model. In the nonlinear model, the marginal productivity of investments is affected by the level of skills established by previous investments. The higher the current level of skills, the higher the productivity of investments. Policies that wait until adolescence to remediate

are more expensive because they attempt to increase investments when complementary stocks of skills from early investments are very low. The lower stock of skills makes the marginal productivity of late remediation investments for disadvantaged children lower.

Skills do not depreciate as quickly in our estimated nonlinear model as they do in the linear model when skills are high and investments are suddenly reduced. Recall the analysis summarized in Figure 11. A linear model implies a much faster rate of depreciation than is found in the data and is produced from the nonlinear model. Nonlinearity reinforces the role of self-productivity in reproducing skills. This is a manifestation of the phenomenon of complementarity and self productivity. Skills beget skills.

The importance of nonlinearity and, consequently, the fact that the marginal productivity of investments depends on the level of skills produced by previous investments generates an important insight for the design of policies. Policies that are balanced increase returns and are more productive. The returns to later investments are greater if high early investments are made. Perry and Abecedarian children made less use of special education than peers who did not receive treatment. The intervention made later schooling more effective. If early interventions are followed up with later interventions, the outcomes can be considerably improved.

We now show the benefits of a balanced intervention in a different way. Table 21 shows the same baseline outcomes that were presented in Table 20 (column one) and the effects of the same Perry-like investments (column two). Column three shows the effect of adding the adolescent-only intervention to the early Perry intervention. There is no counterpart to this intervention in Table 20. Column four shows the outcomes of a Balanced Intervention, which is defined as the constant flow of investment expenditures with the same present value of costs as the intervention reported in column three. While the Adolescent Intervention only increases investments during adolescence, the Balanced Intervention reallocates some of the investments to earlier ages. More specifically, let  $I_0$  denote the level of investments if no intervention is made. The flow of investments in the Adolescent Intervention is  $I_0$  units during the first period from ages 6/7 to ages 8/9, and the second period from ages 8/9 to ages 10/11, but  $I_A$  units during the period from ages 10/11 to ages 12/13. If  $r$  is the interest rate for each period, the present value of investment units of the Adolescent Intervention,  $PV$ , is

$$PV = I_0 + \frac{I_0}{(1+r)} + \frac{I_A}{(1+r)^2}.$$

The “Balanced Intervention” provides a constant flow of investments,  $I_B$ , in all transitions. The level of  $I_B$  is defined according to the equation:

$$PV = I_B + \frac{I_B}{(1+r)} + \frac{I_B}{(1+r)^2}.$$

The Adolescent Intervention coupled with the Perry-like intervention raises the educational attainment of the disadvantaged population by increasing high-school graduation rates and college enrollment to 84% and 27%, respectively. It reduces participation in crime by greatly reducing conviction and probation rates. It is effective in alleviating welfare usage in early adult years.

A balanced policy works even better. For such a program, high school graduation and college enrollment rates are, respectively, 91% and 37%. The reduction in conviction and probation rates is marginally better, and welfare use is reduced to a low 2.6% rate.

Again, complementarity implies that early investments are more productive if they are followed up with late investments. And late investments are more productive if they are preceded by early investments. The mechanism that makes the Balanced Intervention more effective has a very simple economic interpretation. When Adolescent Interventions are made, baseline skills are low and, consequently, so is the marginal productivity of later investments. A balanced investment program increases the stock of skills at the beginning of adolescence. But because the marginal productivity of later investments depends on the level of skills acquired prior to adolescence, the investment in the last period is more productive. Thus, the same amount of total investment distributed more evenly over the life cycle of the child can produce even more adult skill.

We demonstrate this point with another example. As previously noted, the Perry Preschool program increased investments during the program years to around the 7th decile in the distribution of investments. Suppose that the investments in Perry are redistributed so that the same amount of investment is made at each age 0 to age 17 and the present value of investments under this policy is the same as in Perry Preschool program. We call this program the “Perry plus Follow Up Intervention” (PFI). Figure 18 plots the simulated mean cognitive score by age for the Perry control group (solid curve) against the mean cognitive score for the Perry treatment (dashed curve) and the PFI treatment (the dashed-dotted curve). Note that PFI starts at age zero, so by age 3 there is already difference in the level of skills between the

PFI treatment and the entire Perry program.<sup>60</sup> However, once the Perry intervention starts at ages 3/4, the average scores in cognitive skills for the treatment group rises sharply (which is consistent with the Perry data) because of the increase in investments. At age five, the children in Perry would have average cognitive scores more than 0.6 standard deviation above the children in PFI. After the Perry intervention is finished, cognitive scores start to decrease for treatment group members, but they keep increasing for the children in PFI, as investments are kept at a higher level than for the later periods of the Perry-like treatment where no intervention is made. At age 17, the children in the PFI have scores on average 0.5 standard deviations above the Perry treatment children. Again, early investments are specially productive if they are followed up by high later investments.

It is instructive to conduct the counterfactual experiment in which the Perry preschool program participants would receive the same level of investments from age 3 to age 17, instead of only receiving treatment from age 3 to age 5. We call the more intensive program a “Lifetime Perry Program”. Obviously, this program is more expensive, as we are not keeping the present value the same as that in the Perry Program. The effects of this program on cognitive skills are large. Figure 19 plots the mean cognitive score by age for the Perry control group (the solid curve) along with that for the Perry treatment group (the dashed curve) and the Lifetime Perry Program treatment group (the dashed curve). The constant flow of high investments during the childhood and youth years cause the disadvantaged group to score almost half a standard deviation above the mean in cognitive scores. This reinforces the finding of a large literature that finds that remediation is possible if it starts early and if proper follow-up in investment is carried out.

## Part VI

# Conclusion

This paper reviews the evidence on the life cycle of human skill formation. It interprets the evidence using basic economic models. It estimates the economic models using rich data on measurements of skills of children and parents as well as parental investments in the skills of the children. The estimation results

---

<sup>60</sup>We remind the reader that the Perry program was a randomized experiment and at the age of entry there was no significant differences in cognitive skills between control and treatment groups.

provides further evidence that the childhood is a multistage process where early investments feed into later investments. Skill begets skill; learning begets learning. Research in economics collapses childhood into a single period and implicitly assumes that all investments at all ages of the child are perfect substitutes. This misses important features of the skill development process.

The evidence reported here is broadly consistent with the self-productivity of human capital investment and the complementarity of investments at different ages. Both factors combine to produce the phenomenon that skill begets skill. Complementarity implies that early investments need to be followed by later investments if the early investments are to pay off.

This paper formalizes the concept of critical and sensitive periods and introduces the concepts of complementarity and self-productivity on the child development process. Complementarity and self-productivity produce no trade-off between equity and efficiency at early ages of human development but a substantial trade-off at later ages. Once skills are crystallized, complementarity implies that the returns are highest for investment in the most able. At the youngest ages, it is possible to form ability and create the complementarity that characterizes late adolescent and early adult human capital investment processes. Thus early interventions targeted toward the disadvantaged can be highly effective if they are followed up with later investments. Similarly, late interventions only produce substantial results if they are anticipated with high investments as well.

The main findings of the literature and our new empirical estimates can be summarized as follows. First, abilities matter. A large number of empirical studies document that cognitive ability affects both the likelihood of acquiring advanced training and higher education, and the economic returns to those activities. Both cognitive and noncognitive abilities matter in determining participation in crime, teenage pregnancy, drug use and participation in other deviant activities. The evidence that abilities matter tells us nothing whatsoever about whether they are genetically determined.

Second, ability is multidimensional. IQ has to be distinguished from what is measured by achievement tests, although it partly determines success on achievement tests. Noncognitive skills (perseverance, motivation, self-control and the like) have direct effects on wages (given schooling), schooling, teenage pregnancy, smoking, crime and achievement tests. Both cognitive and noncognitive skills affect socio-economic success. Both are strongly influenced by family environments. The old dichotomy between an invariant, genetically determined ability and acquired skills is a false one that still continues to influence

the literature in economics. Abilities and skills are both acquired. They are influenced both by genes and the environment.

Third, ability gaps in both cognitive and noncognitive skills across individuals and across socioeconomic groups open up at early ages. They are strongly correlated with family background factors, like parental education and maternal ability, which, when controlled for in a statistical sense, largely eliminate these gaps. Inputs of schooling quality and resources have relatively small effects on early ability deficits. Parenting practices have strong effects on emotional development and motivation. This correlational evidence is supported by the experimental evidence from the Perry Preschool Program and the Abecedarian program.

Fourth, it is possible to partially compensate for adverse family environments. Evidence from randomized trials conducted on intervention programs targeted at disadvantaged children who are followed into adulthood, suggests that it is possible to eliminate some of the gaps due to early disadvantage. Enriched and sustained interventions at the youngest ages raise IQ. The Abecedarian program provided an enriched intervention for disadvantaged children starting at age 4 months. The children who received the intervention score consistently higher than the children who do not, even long after the treatment is discontinued. Later interventions like the Perry Preschool program show no lasting effect on IQ. However, effects on motivation and, hence, achievement test scores are found. Children are less likely to commit crime and have out of wedlock births and are more likely to participate in regular schooling. Early interventions have a substantial effect on adult performance and have a high economic return.

Fifth, different types of abilities appear to be manipulable at different ages. Thus, while factors affecting IQ deficits need to be addressed at very early ages for interventions to be effective, there is evidence that later interventions in the adolescent years can affect noncognitive skills as well as the knowledge measured by achievement tests. Achievement is determined by both cognitive and noncognitive factors. This evidence is rooted in the neuroscience that establishes the malleability of the prefrontal cortex into the early 20s. This is the region of the brain that governs emotion and self-regulation.

Sixth, the later the remediation process starts, the less effective it is. Classroom remediation programs designed to combat early cognitive deficits have a poor track record. Public job training programs and adult literacy and educational programs, like the GED, that attempt to remediate years of educational and emotional neglect among disadvantaged individuals have a low economic return, and for young males, the return is negative. This evidence is consistent with strong complementarity of investment over the life

cycle of the child.

Seventh, the economic returns to initial investments at early ages are high. The economic return to investment at older ages is lower. The technology of skill formation which we analyze in this essay suggests a strong skill multiplier effect of investment. Investment at an early age produces a high return through self-productivity and direct complementarity. Early investment in cognitive and noncognitive skills lowers the cost of later investment by making learning at later ages more efficient. The skill multiplier highlights the value of early investment. It also demonstrates that there is no trade-off between equity (targeting programs at disadvantaged families) and efficiency (getting the highest economic returns), provided that the investments are made at early ages. There is such a trade-off at later ages.

Eighth, CES-complementarity of early with late investments implies that early investments must be followed up by later investments in order to be effective. Nothing in the new economics of human skill formation suggests that we should starve later educational and skill enhancement efforts. Our evidence suggests that a portfolio of childhood investments tipped towards the younger years of a child's life is optimal. However, we should prioritize, and shift our priorities, in a marginal fashion by redirecting a given total sum of expenditure on skill investment to earlier ages relative to how it is currently allocated toward disadvantaged populations that do not provide enriched environments for their children.

## References

- Aiyagari, S. R., J. Greenwood, and A. Seshadri (2002, February). Efficient investment in children. *Journal of Economic Theory* 102(2), 290–321.
- Altonji, J. G., T. E. Elder, and C. R. Taber (2005, February). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy* 113(1), 151–184.
- Anderson, D. A. (1999, October). The aggregate burden of crime. *Journal of Law and Economics* 42(2), 611–642.
- Armor, D. J. (2003). *Maximizing Intelligence*. New Brunswick, NJ: Transaction Publishers.
- Barnett, W. S. and L. N. Masse (2002). A benefit-cost analysis of the abecedarian early childhood intervention. Technical report, National Institute for Early Education Research (NIEER), New Brunswick, New Jersey.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research, distributed by Columbia University Press.
- Becker, G. S. and N. Tomes (1979, December). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* 87(6), 1153–1189.
- Becker, G. S. and N. Tomes (1986, July). Human capital and the rise and fall of families. *Journal of Labor Economics* 4(3, Part 2), S1–S39.
- Belfield, C. R., M. Nores, and W. S. Barnett (2004). The High/Scope Perry Pre-School Program: Cost–benefit analysis using data from the age-40 follow-up. Technical report, National Institute for Early Education Research (NIEER), New Brunswick, New Jersey.
- Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75(4, Part 1), 352–365.
- Benabou, R. (2000, March). Unequal societies: Income distribution and the social contract. *American Economic Review* 90(1), 96–129.

- Benabou, R. (2002, March). Tax and education policy in a heterogeneous agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica* 70(2), 481–517.
- Bowles, S. and H. Gintis (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. New York: Basic Books.
- Bradley, R. H. and B. M. Caldwell (1980, December). The relation of home environment, cognitive competence, and IQ among males and females. *Child Development* 51(4), 1140–1148.
- Bradley, R. H. and B. M. Caldwell (1984, June). The relation of infants' home environments to achievement test performance in first grade: A follow-up study. *Child Development* 55(3), 803–809.
- Brooks-Gunn, J., L. J. Berlin, and A. S. Fuligni (2000). Early childhood intervention programs: What about the family? In S. Meisels and J. Shonkoff (Eds.), *The Handbook of Early Childhood Interventions*, pp. 549–588. New York: Cambridge University Press.
- Brooks-Gunn, J., R. Gross, H. Kraemer, D. Spiker, and S. Shapiro (1992, June). Enhancing the cognitive outcomes of low birth weight, premature infants: For whom is the intervention most effective? *Pediatrics* 89(6, Part 2), 1209–1215.
- Bureau of Justice (2004). Bureau of justice statistics criminal offenders statistics. Website.
- Cameron, J. (2004). Evidence for an early sensitive period for the development of brain systems underlying social affiliative behavior. Unpublished manuscript, Oregon National Primate Research Center.
- Cameron, S. V. and J. J. Heckman (1993, January). The nonequivalence of high school equivalents. *Journal of Labor Economics* 11(1, Part 1), 1–47.
- Cameron, S. V. and J. J. Heckman (2001, June). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy* 109(3), 455–99.
- Cardak, B. A. and D. Givon (2004, August). Why the poor don't go to university: Attainment constraints and two-staged education. Presented at the Fifth APET International Conference on Public Economics, Peking University (Beijing, China), August 25th-29th, 2004.

- Carneiro, P. and J. J. Heckman (2003). Human capital policy. In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.
- Carneiro, P., J. J. Heckman, and D. V. Masterov (2005, April). Labor market discrimination and racial differences in pre-market factors. *Journal of Law and Economics* 48(1), 1–39.
- Caucutt, E. and K. Kumar (2003, June). Higher education subsidies and heterogeneity: A dynamic analysis. *Journal of Economic Dynamics and Control* 27(8), 1459–1502.
- Caucutt, E. and L. J. Lochner (2004). Early and late human capital investments, credit constraints, and the family. Unpublished manuscript, University of Western Ontario, Department of Economics.
- Cunha, F. (2004). Skill formation in a laiter economy. Ph.D. thesis, University of Chicago.
- Cunha, F. and J. J. Heckman (2006). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Unpublished manuscript, University of Chicago, Department of Economics. Forthcoming, *Journal of Human Resources*.
- 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*. Amsterdam: North-Holland. Forthcoming.
- Cunha, F., J. J. Heckman, and S. M. Schennach (2006). Estimating the technology of cognitive and noncognitive skill formation. Unpublished manuscript, University of Chicago, Department of Economics. Presented at the Yale Conference on Macro and Labor Economics, May 5–7, 2006.
- Currie, J. (2001, Spring). Early childhood education programs. *Journal of Economic Perspectives* 15(2), 213–238.
- Currie, J. and D. Blau (2006). Who’s minding the kids? Preschool, day care, and after school care. In F. Welch and E. Hanushek (Eds.), *Handbook of the Economics of Education*. Amsterdam: North-Holland. forthcoming.

- Currie, J. and D. Thomas (2000, Fall). School quality and the longer-term effects of Head Start. *Journal of Human Resources* 35(4), 755–774.
- DeLong, J. B., L. Katz, and C. Goldin (2003). Sustaining american economic growth. In H. Aaron, J. Lindsay, and P. Nivola (Eds.), *Agenda for the Nation*. Washington, DC: Brookings Institution Press.
- Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. New York: Springer-Verlag.
- Duncan, G., A. Claessens, and M. Engel (2004). The contributions of hard skills and socio-emotional behavior to school readiness. Unpublished manuscript, Northwestern University.
- Duncan, G., A. Kalil, S. E. Mayer, R. Tepper, and M. R. Payne (2005). The apple does not fall far from the tree. In S. Bowles, H. Gintis, and M. Osborne Groves (Eds.), *Unequal Chances: Family Background and Economic Success*. Princeton: Princeton University Press.
- Duncan, G. J. and J. Brooks-Gunn (1997). Income effects across the life span: Integration and interpretation. In G. Duncan and J. Brooks-Gunn (Eds.), *Consequences of Growing Up Poor*, pp. 596–610. New York: Russell Sage Foundation.
- Edwards, R. C. (1976, Winter). Individual traits and organizational incentives: What makes a “good” worker? *Journal of Human Resources* 11(1), 51–68.
- Ellwood, D. T. (2001). The sputtering labor force of the twenty-first century: Can social policy help? In A. Krueger and R. Solow (Eds.), *The Roaring Nineties: Can Full Employment Be Sustained?*, pp. 421–489. New York: Russell Sage Foundation.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* 101, 171–191.
- Fuerst, J. and D. Fuerst (1993, April). Chicago experience with an early childhood program: The special case of the child parent center program. *Urban Education* 28(1), 69–96.
- Garces, E., D. Thomas, and J. Currie (2002, September). Longer-term effects of Head Start. *American Economic Review* 92(4), 999–1012.

- Gomby, D. S., C. S. Larson, E. M. Lewit, and R. E. Behrman (1999, Spring/Summer). Home visiting: Recent program evaluations—analysis and recommendations. *The Future of Children* 9(1), 4–26.
- Gormley, W., T. Gayer, D. Phillips, and B. Dawson (2004). The effect of universal pre-k on cognitive development. Unpublished manuscript, Georgetown University.
- Griliches, Z. (1977, January). Estimating the returns to schooling: Some econometric problems. *Econometrica* 45(1), 1–22.
- Grogger, J. and D. Neal (2000). Further evidence of the effects of catholic secondary schooling. *Brookings-Wharton Papers on Urban Affairs*, 151–193.
- Hammersley, J. M. and K. W. Morton (1954). Poor man’s Monte Carlo. *Journal of the Royal Statistical Society. Series B (Methodological)* 16(1), 23–38.
- Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004, July–August). The effect of schooling and ability on achievement test scores. *Journal of Econometrics* 121(1-2), 39–98.
- Hanushek, E., C. K. Y. Leung, and K. Yilmaz (2004). Borrowing constraints, college aid, and intergenerational mobility. Technical Report 10711, NBER.
- Heckman, J. J. (1995, October). Lessons from *The Bell Curve*. *Journal of Political Economy* 103(5), 1091.
- Heckman, J. J. (1999, Spring). Doing it right: Job training and education. *The Public Interest* (135), 86–107.
- Heckman, J. J. (2000, March). Policies to foster human capital. *Research in Economics* 54(1), 3–56. With discussion.
- Heckman, J. J. (2004). Lessons from the technology of skill formation. *Annals of the New York Academy of Science* (1038), 179–200.
- Heckman, J. J. and P. LaFontaine (2006, July). Bias corrected estimates of GED returns. *Journal of Labor Economics* 24(3), 661–700.

- Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 31, pp. 1865–2097. New York: North-Holland.
- Heckman, J. J., M. I. Larenas, and S. Urzua (2004). Accounting for the effect of schooling and abilities in the analysis of racial and ethnic disparities in achievement test scores. Unpublished manuscript, University of Chicago, Department of Economics.
- Heckman, J. J. and L. J. Lochner (2000). Rethinking myths about education and training: Understanding the sources of skill formation in a modern economy. In S. Danziger and J. Waldfogel (Eds.), *Securing the Future: Investing in Children from Birth to College*. New York: Russell Sage Foundation.
- Heckman, J. J., L. J. Lochner, and C. Taber (1998, January). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1(1), 1–58.
- Heckman, J. J., L. J. Lochner, and P. E. Todd (2004). Earnings functions and rates of return: the mincer equation and beyond. Unpublished manuscript, University of Chicago, Department of Economics.
- Heckman, J. J. and Y. Rubinstein (2001, May). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review* 91(2), 145–149.
- 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.
- International Adult Literacy Survey (2002). *International Adult Literacy Survey Microdata User's Guide*. Ottawa: Statistics Canada.
- Johnson, A. W. (1996). *An Evaluation of the Long-Term Impacts of the Sponsor-A-Scholar Program on Student Performance*. Princeton, NJ: Mathematica Policy Research.
- Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2003). Growth of US industries and investments in information technology and higher education. In C. Corrado, J. C. Haltiwanger, and D. E. Sichel (Eds.), *Measuring Capital in the New Economy*. Chicago, IL: University of Chicago Press.

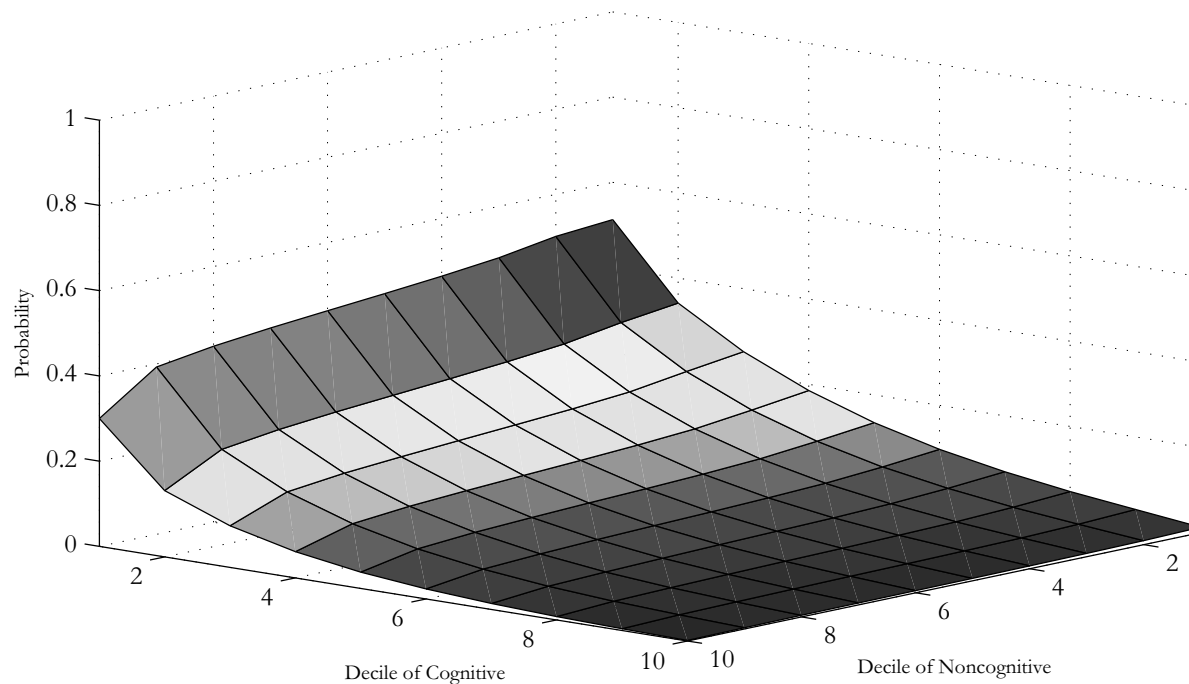
- Karoly, L. A., P. W. Greenwood, S. S. Everingham, J. Hoube, M. R. Kilburn, C. P. Rydell, M. Sanders, and J. Chiesa (1998). *Investing in our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*. Santa Monica, CA: RAND.
- Klein, R., R. Spady, and A. Weiss (1991, October). Factors affecting the output and quit propensities of production workers. *Review of Economic Studies* 58(5), 929–953.
- Knudsen, E. I. (2004, October). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience* 16(1), 1412–1425.
- 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.
- Levitt, S. (1997, June). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review* 87(3), 270–290.
- Linver, M. R., J. Brooks-Gunn, and N. Cabrera (2004, April-September). The home observation for measurement of the environment (HOME) inventory: The derivation of conceptually designed subscales. *Parenting: Science & Practice* 4(2/3), 99–114.
- Lochner, L. J. and E. Moretti (2004, March). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review* 94(1), 155–189.
- Love, J. M., E. Eliason-Kisker, C. M. Ross, P. Z. Schochet, J. Brooks-Gunn, D. Paulsell, K. Boller, J. Constantine, C. Vogel, A. Sidle Fuligni, and C. Brady-Smith (2002). Making a difference in the lives of infants and toddlers and their families: The impacts of early Head Start. Technical Report PR02-30a, Mathematica Policy Research Report.
- Magnuson, K. A., M. Meyers, C. J. Ruhm, and J. Waldfogel (2004, Spring). Inequality in preschool education and school readiness. *American Educational Research Journal* 41(1), 115–157.
- Magnuson, K. A., C. J. Ruhm, and J. Waldfogel (2004). Does prekindergarten improve school preparation and performance? Technical Report 10452, NBER.

- Maxfield, M., A. Schirm, and N. Rodriguez-Planas (2003, August). The quantum opportunity program demonstration: Implementation and short-term impacts. Technical Report MPR Reference No.: 8279-093, Mathematica Policy Research, Inc., Washington, DC.
- Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* 24(1), 1161–1192.
- Miller, T. R., M. A. Cohen, and B. Wiersema (1996). *Victim Costs and Consequences: A New Look*. Washington, D.C.: U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice.
- Morris, P., G. J. Duncan, and E. Clark-Kauffman (2005). Child well-being in an era of welfare reform: The sensitivity of transitions in development to policy change. *Developmental Psychology* 41(6), 919–932.
- Moskoff, W. (1980, April). Pronatalist policies in romania. *Economic Development and Cultural Change* 28(3), 597–614.
- Newport, E. L. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*, pp. 737–740. London: Macmillan Publishers Ltd./Nature Publishing Group.
- O'Connor, T. G., M. Rutter, C. Beckett, L. Keaveney, J. M. Kreppner, the English, and Romanian Adoptees Study Team (2000, March-April). The effects of global severe privation on cognitive competence: Extension and longitudinal follow-up. *Child Development* 71(2), 376–390.
- Olds, D. L. (2002, September). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science* 3(2), 153–172.
- Peterson, J. L. and N. Zill (1986, May). Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family* 48(2), 295–307.
- Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: W. Morrow and Co.
- Reynolds, A. J., S.-R. Ou, and J. W. Topitzes (2004, September–October). Paths of effects of early childhood interventions on educational attainment and delinquency: A confirmatory analysis of the chicago parent-child centers. *Child Development* 75(5), 1299–1328.

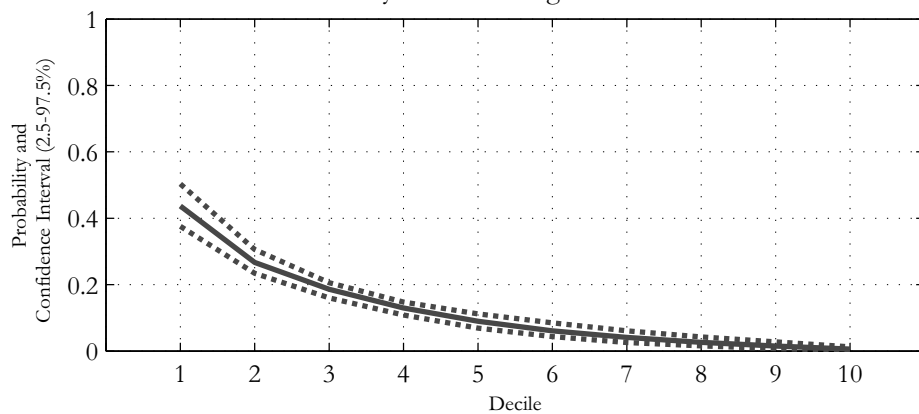
- Rolnick, A. and R. Grunewald (2003). Early childhood development: Economic development with a high public return. Technical report, Federal Reserve Bank of Minneapolis, Minneapolis, MN.
- Rosapepe, J. C. (2001). Half way home: Romania's abandoned children ten years after the revolution. Report of the US Ambassador to Romania. accessed on 12/05/2004.
- Rutter, M. and the English and Romanian Adoptees Study Team (1998, May). Developmental catch-up, and deficit, following adoption after severe global early privation. *Journal of Child Psychology and Psychiatry* 39(4), 465–476.
- Schweinhart, L. J., H. V. Barnes, and D. Weikart (1993). *Significant Benefits: The High-Scope Perry Preschool Study Through Age 27*. Ypsilanti, MI: High/Scope Press.
- Seshadri, A. and K. Yuki (2004, October). Equity and efficiency effects of redistributive policies. *Journal of Monetary Economics* 51(7), 1415–1447.
- Shonkoff, J. P. and D. Phillips (2000). *From Neurons to Neighborhoods: The Science of Early Child Development*. Washington, DC: National Academy Press.
- Spence, A. M. (1973, August). Job market signaling. *Quarterly Journal of Economics* 87(3), 355–374.
- Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. New York: Cambridge University Press.
- Stiglitz, J. E. (1975, June). The theory of 'screening,' education, and the distribution of income. *American Economic Review* 65(3), 283–300.
- Su, X. (2004, December). The allocation of public funds in a hierarchical educational system. *Journal of Economic Dynamics and Control* 28(12), 2485–2510.
- Suomi, S. J. (1999). Developmental trajectories, early experiences, and community consequences: Lessons from studies with rhesus monkeys. In D. P. Keating and C. Hertzman (Eds.), *Developmental Health and the Wealth of Nations: Social, Biological, and Educational Dynamics*, pp. 185–200. The Guilford Press.
- Taggart, R. (1995). *Quantum Opportunity Program*. Philadelphia, PA: Opportunities Industrialization Centers of America, Inc.

- Tierney, J. P., J. B. Grossman, and N. L. Resch (1995). *Making a Difference: An Impact Study of Big Brothers/Big Sisters*. Philadelphia, PA: Public/Private Ventures.
- 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 (2005). The production of cognitive achievement in children: Home, school and racial test score gaps. Unpublished manuscript. Under revision for publication.
- Turkheimer, E., A. Haley, M. Waldron, B. D’Onofrio, and I. I. Gottesman (2003, November). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* 14(6), 623–628.

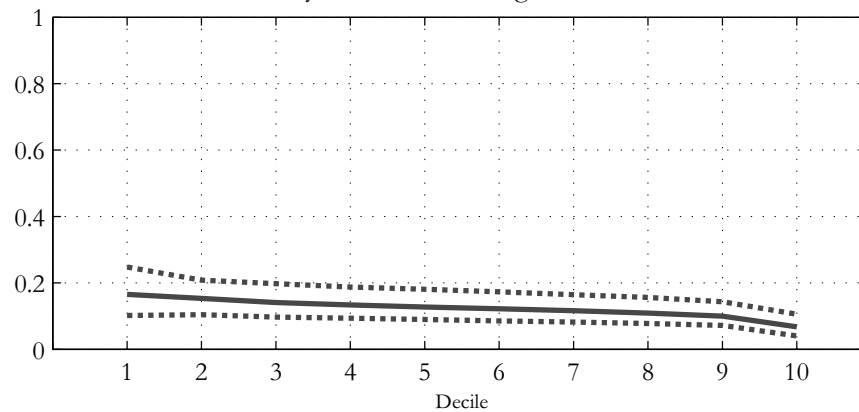
Figure 1A. Probability of Being a High School Dropout by Age 30 - Males  
 i. By Decile of Cognitive and Noncognitive Factors



ii. By Decile of Cognitive Factor

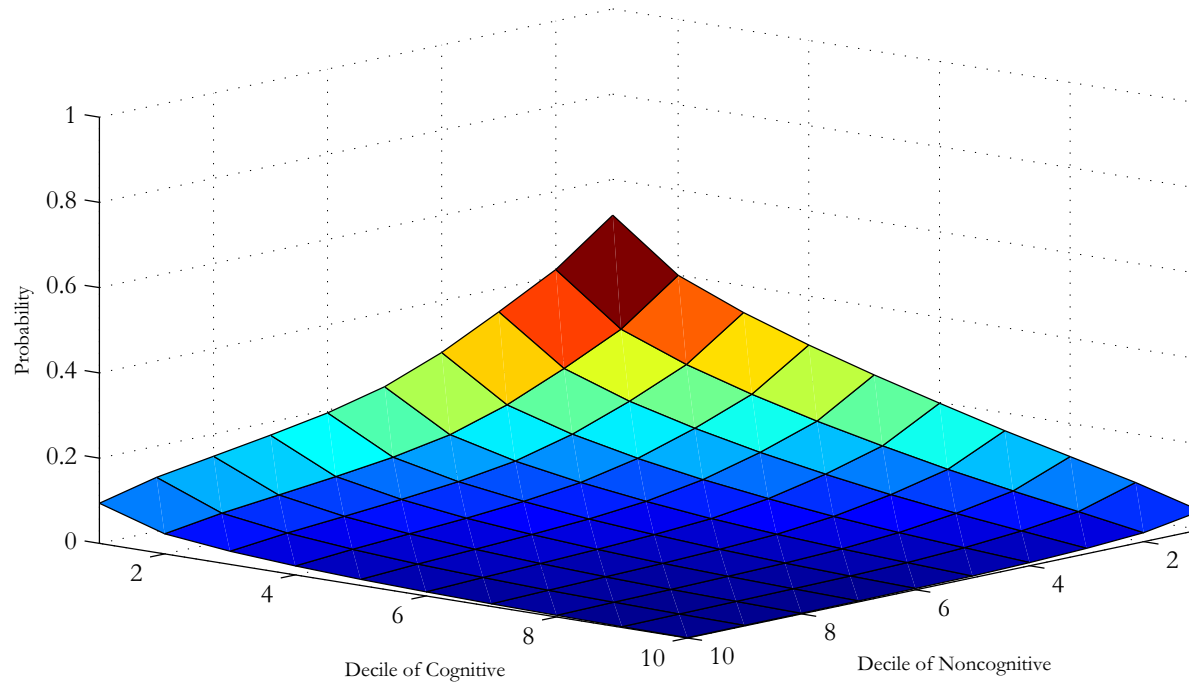


iii. By Decile of Noncognitive Factor

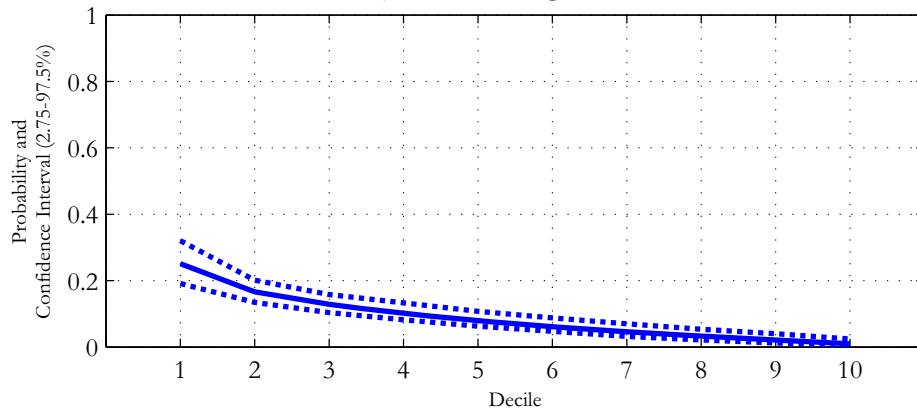


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

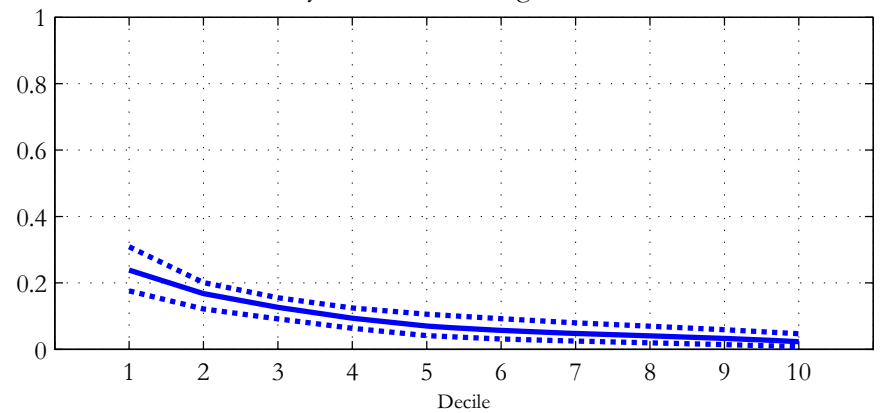
Figure 1B. Probability of Being a High School Dropout by Age 30 - Females  
 i. By Decile of Cognitive and Noncognitive Factors



ii. By Decile of Cognitive Factor

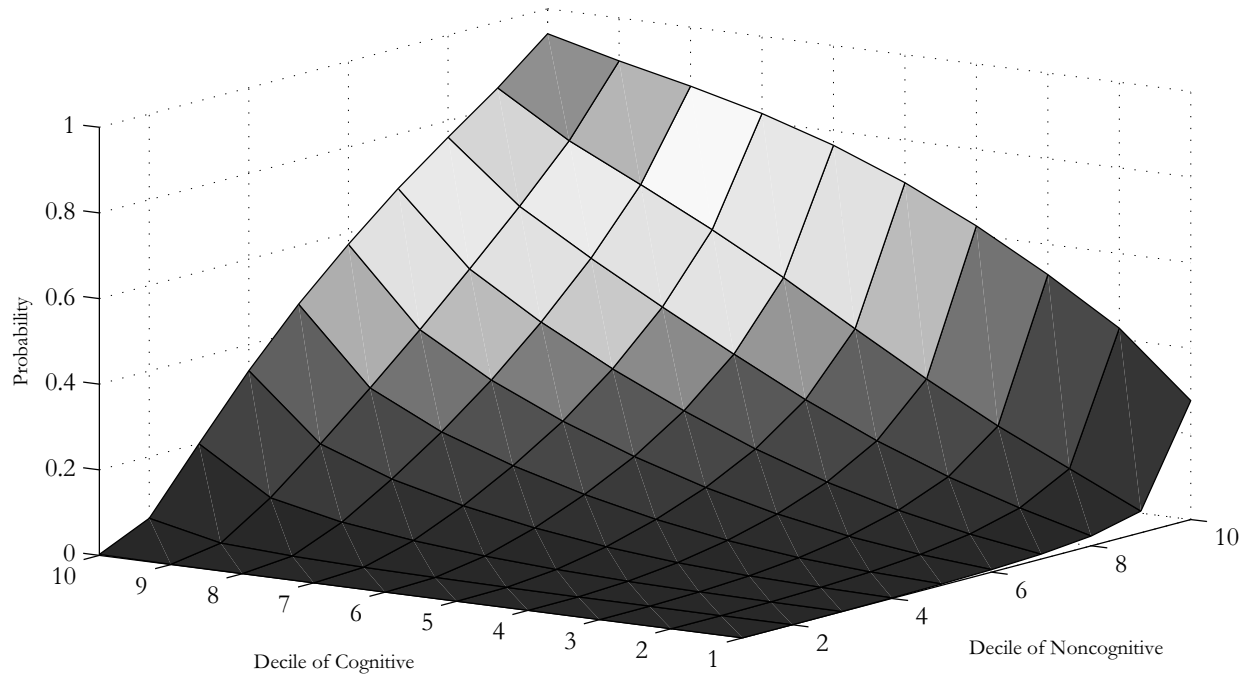


iii. By Decile of Noncognitive Factor

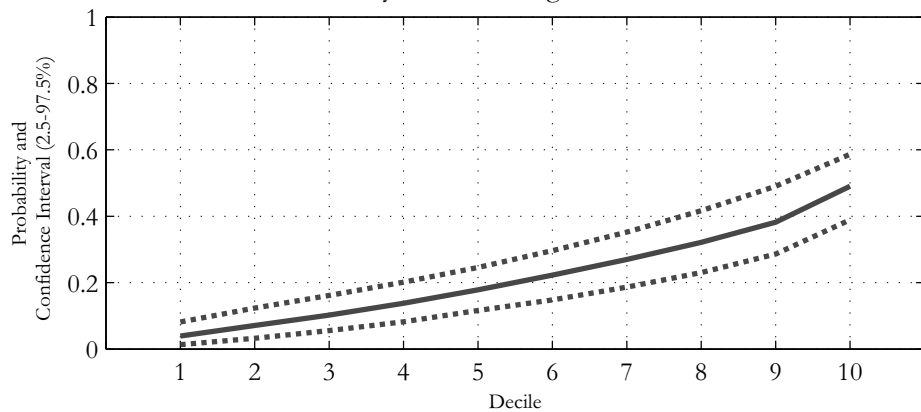


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

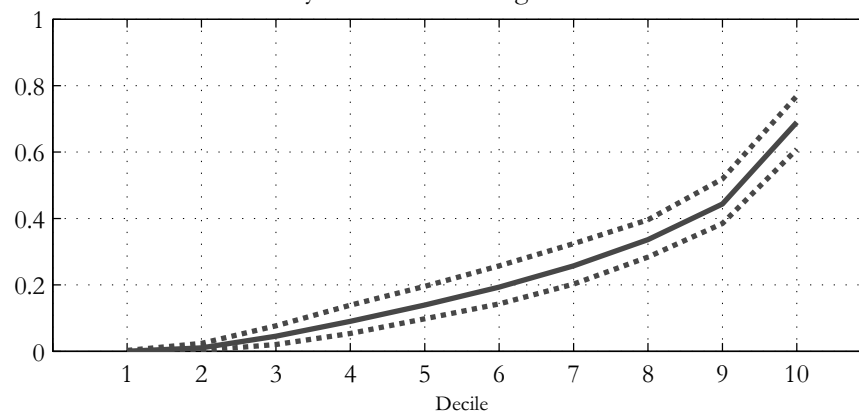
Figure 1C. Probability of Being a 4-yr College Graduate by Age 30 - Males  
 i. By Decile of Cognitive and Noncognitive Factors



ii. By Decile of Cognitive Factor

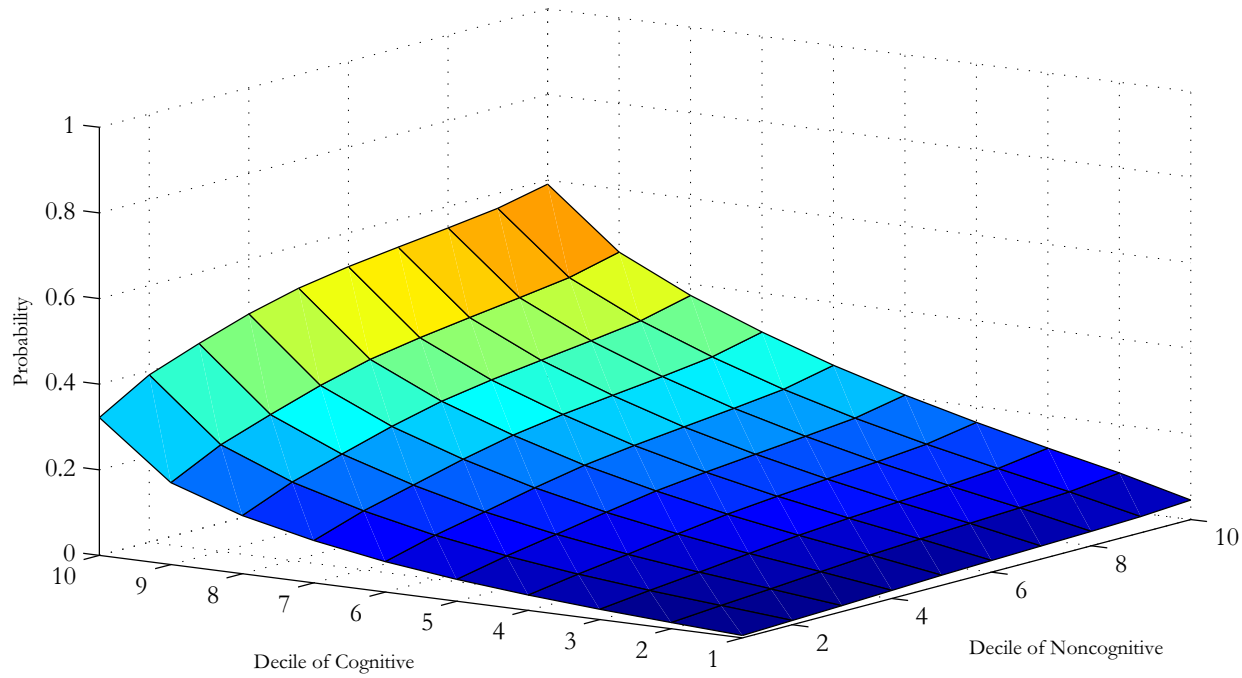


iii. By Decile of Noncognitive Factor

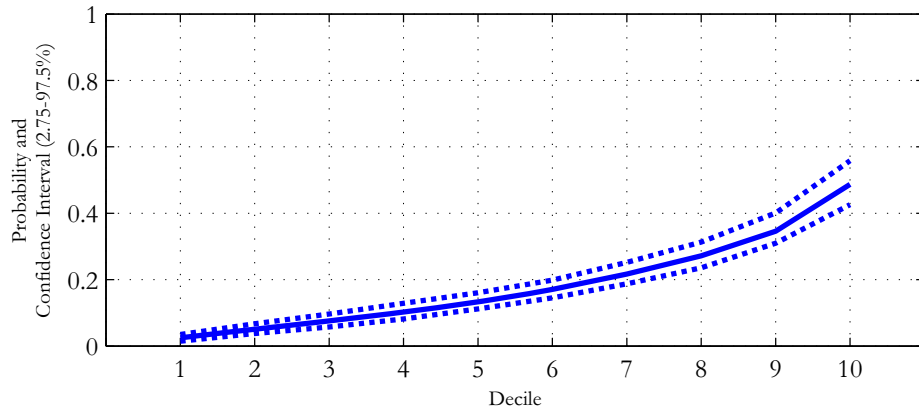


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

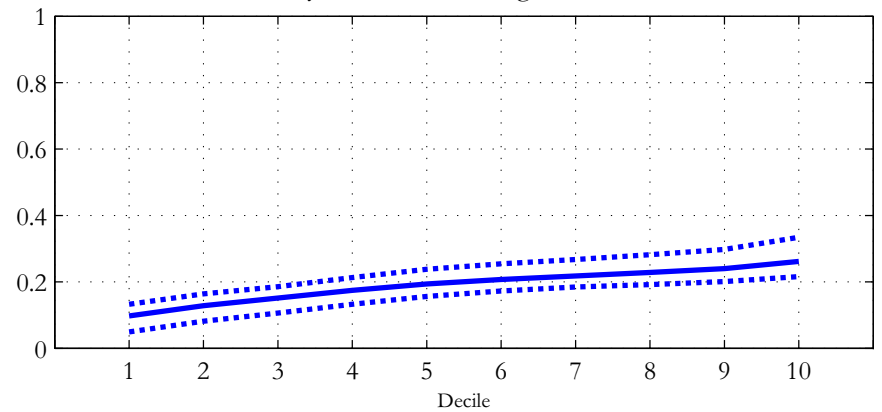
Figure 1D. Probability of Being a 4-yr College Graduate by Age 30 - Females  
 i. By Decile of Cognitive and Noncognitive Factors



ii. By Decile of Cognitive Factor

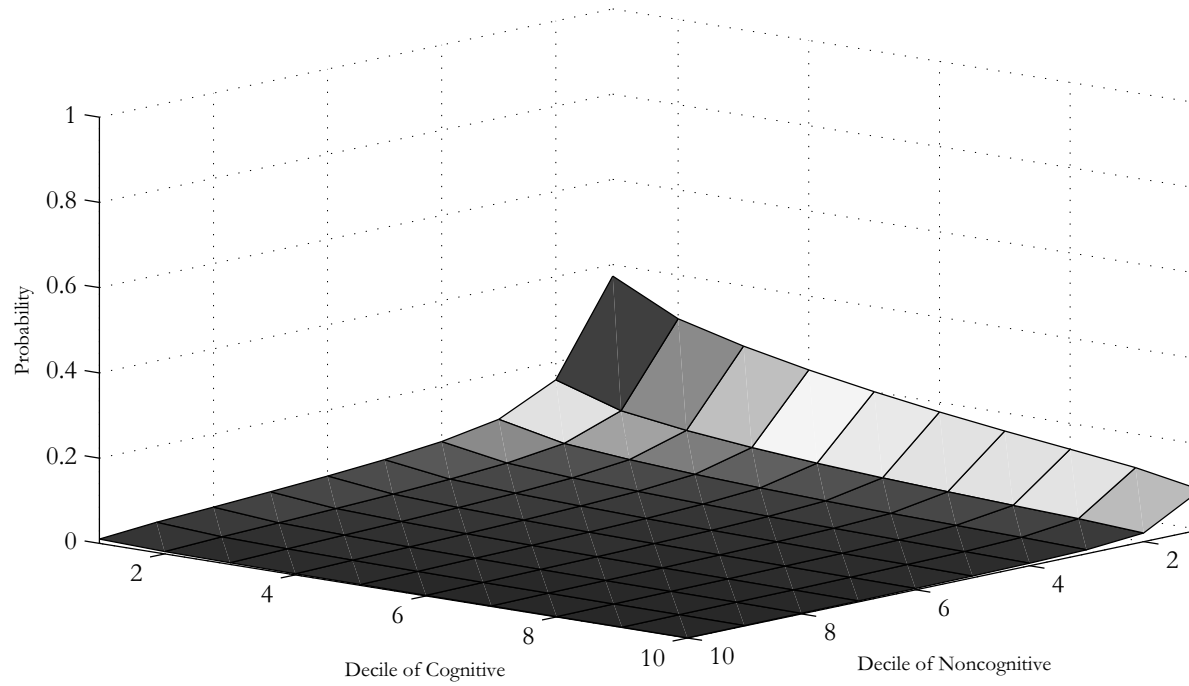


iii. By Decile of Noncognitive Factor

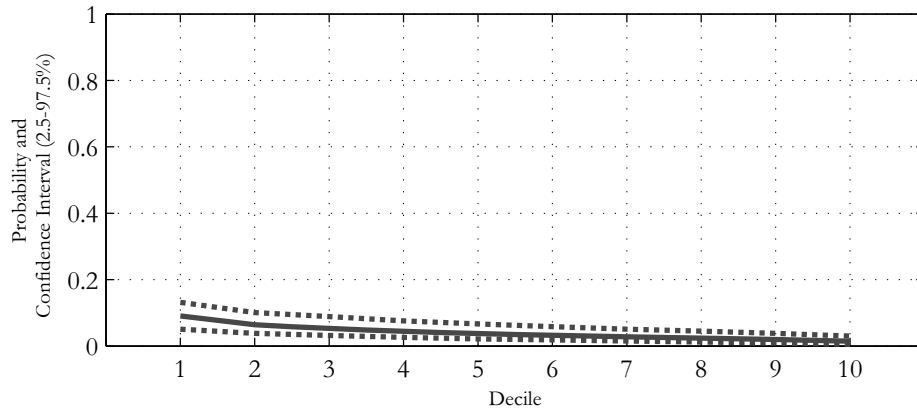


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

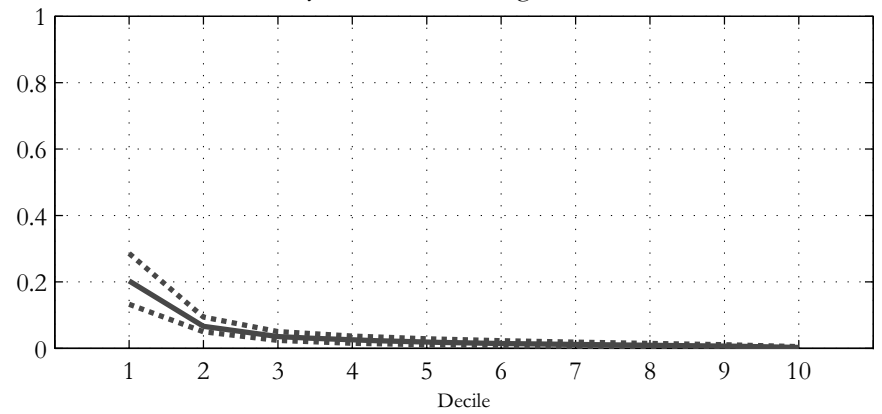
Figure 1E. Probability of Incarceration by Age 30 - Males  
 i. By Decile of Cognitive and Noncognitive Factor



ii. By Decile of Cognitive Factor

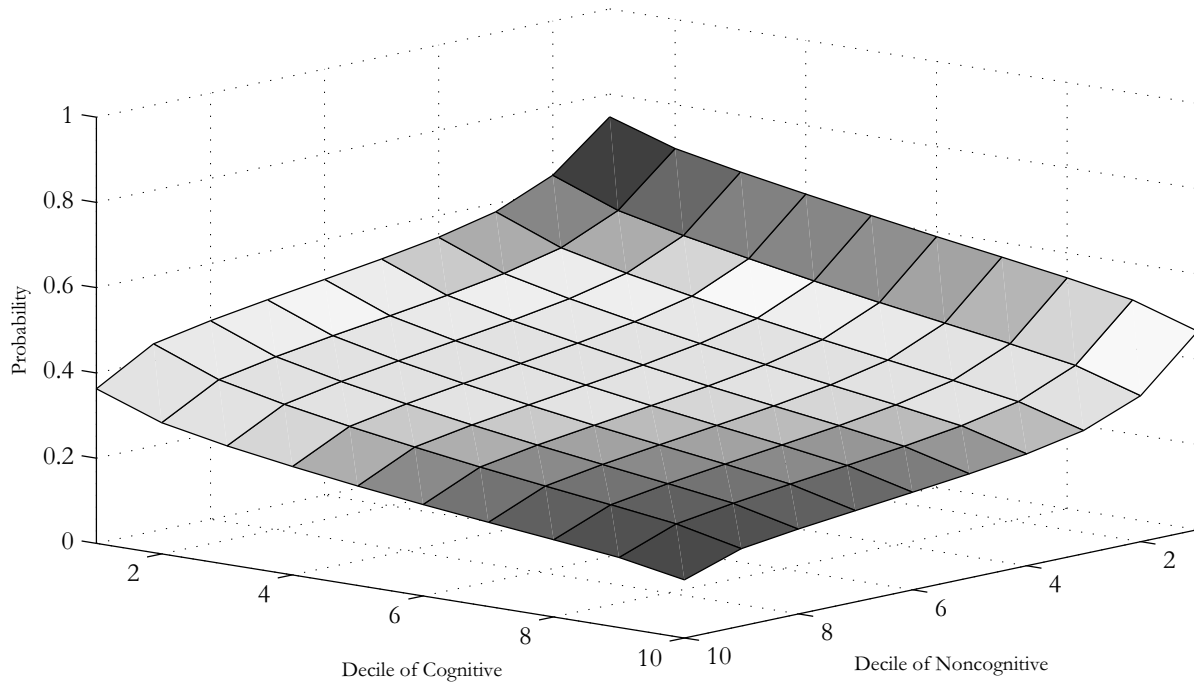


iii. By Decile of Noncognitive Factor

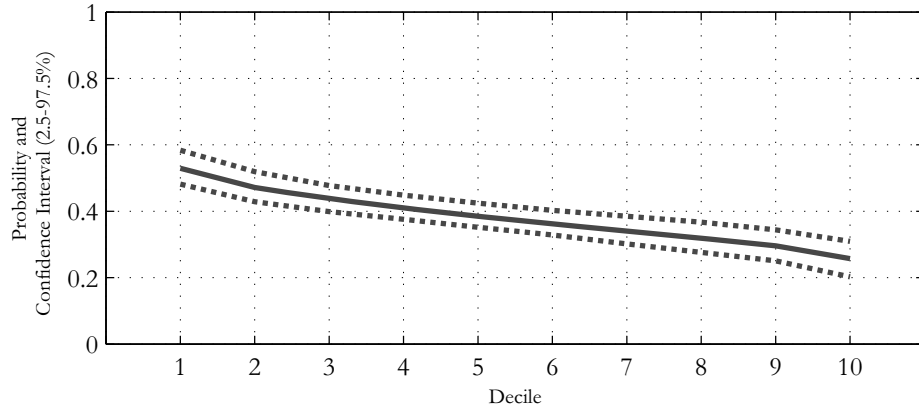


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

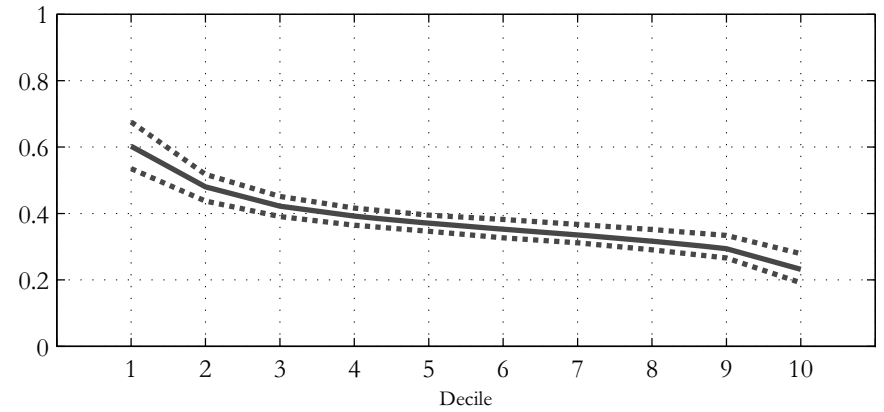
Figure 1F. Probability Of Daily Smoking By Age 18 - Males  
 i. By Decile of Cognitive and Noncognitive Factor



ii. By Decile of Cognitive Factor

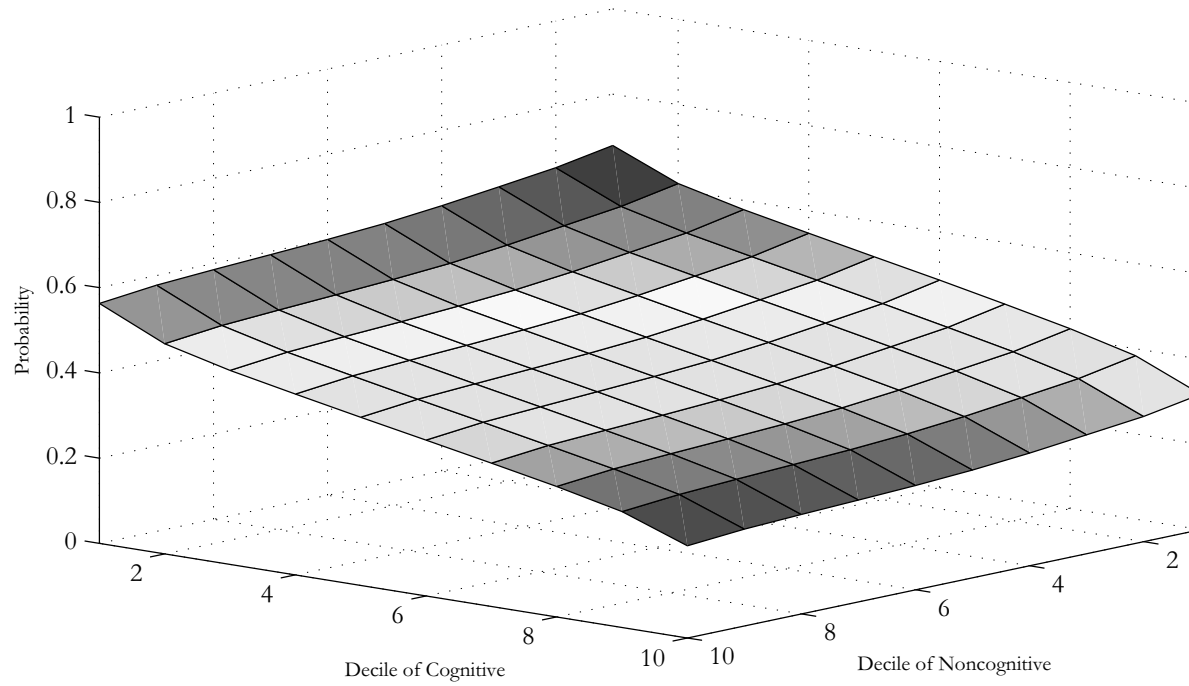


iii. By Decile of Noncognitive Factor

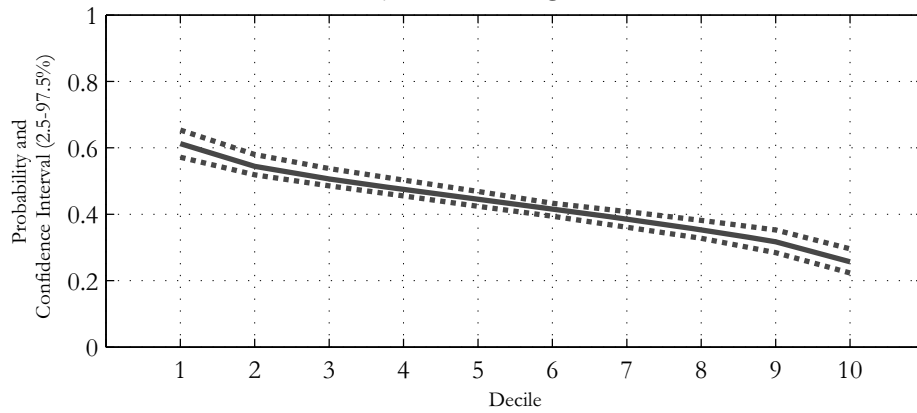


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

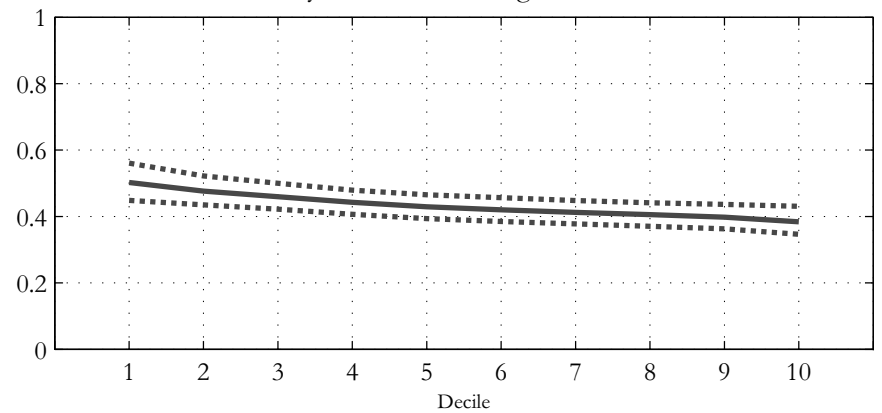
Figure 1G. Probability Of Daily Smoking By Age 18 - Females  
 i. By Decile of Cognitive and Noncognitive Factor



ii. By Decile of Cognitive Factor

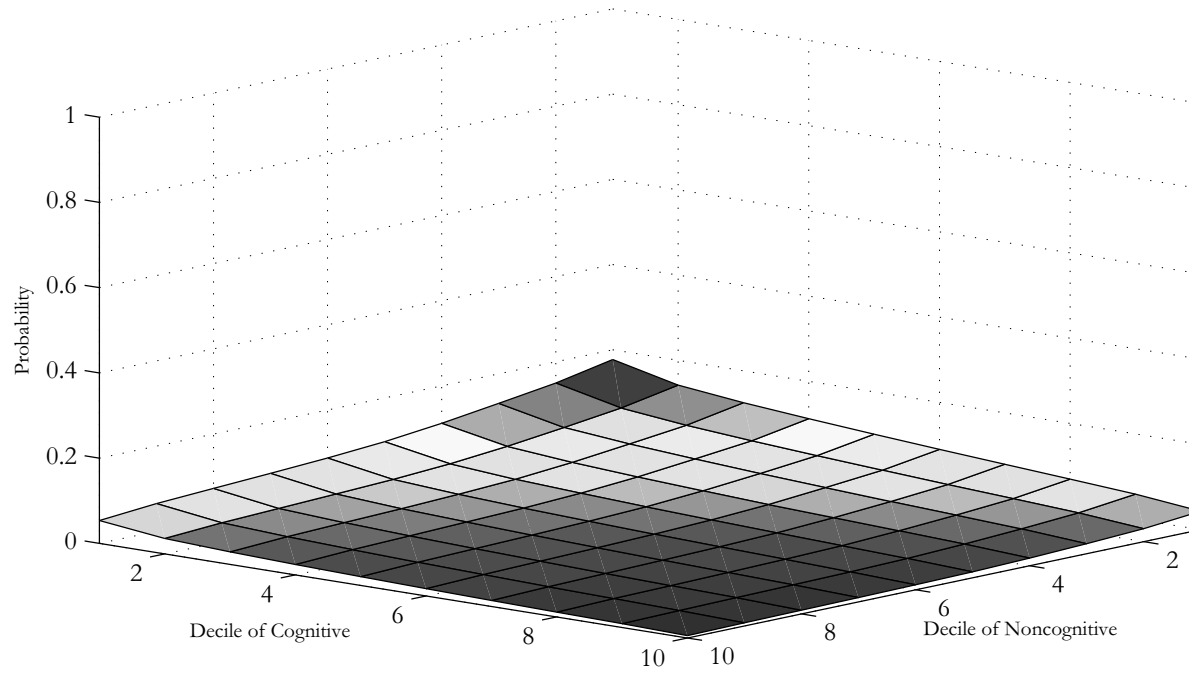


iii. By Decile of Noncognitive Factor

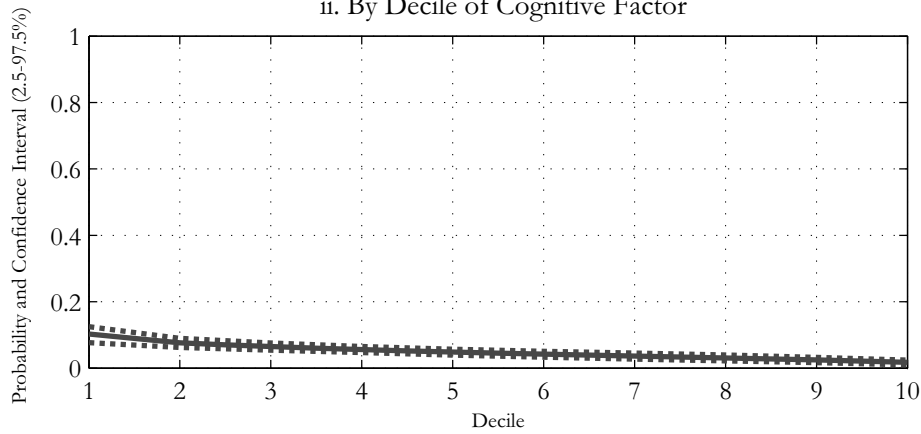


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

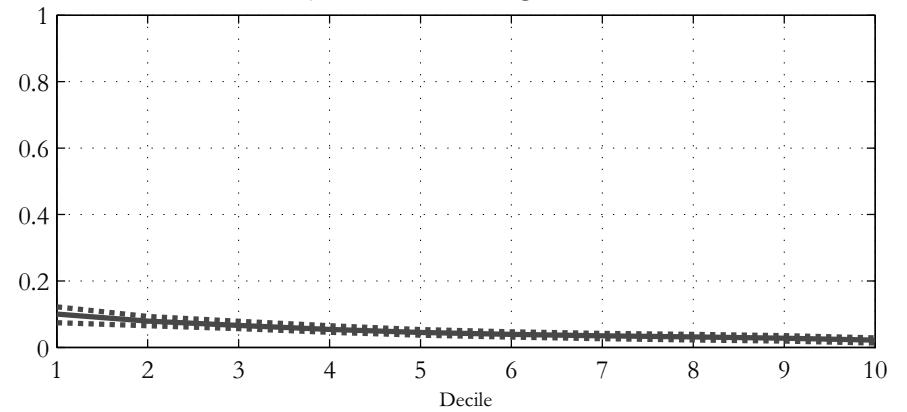
Figure 1H. Probability Of Being Single With Child at Age 18- Females  
 i. By Decile of Cognitive and Noncognitive Factors



ii. By Decile of Cognitive Factor



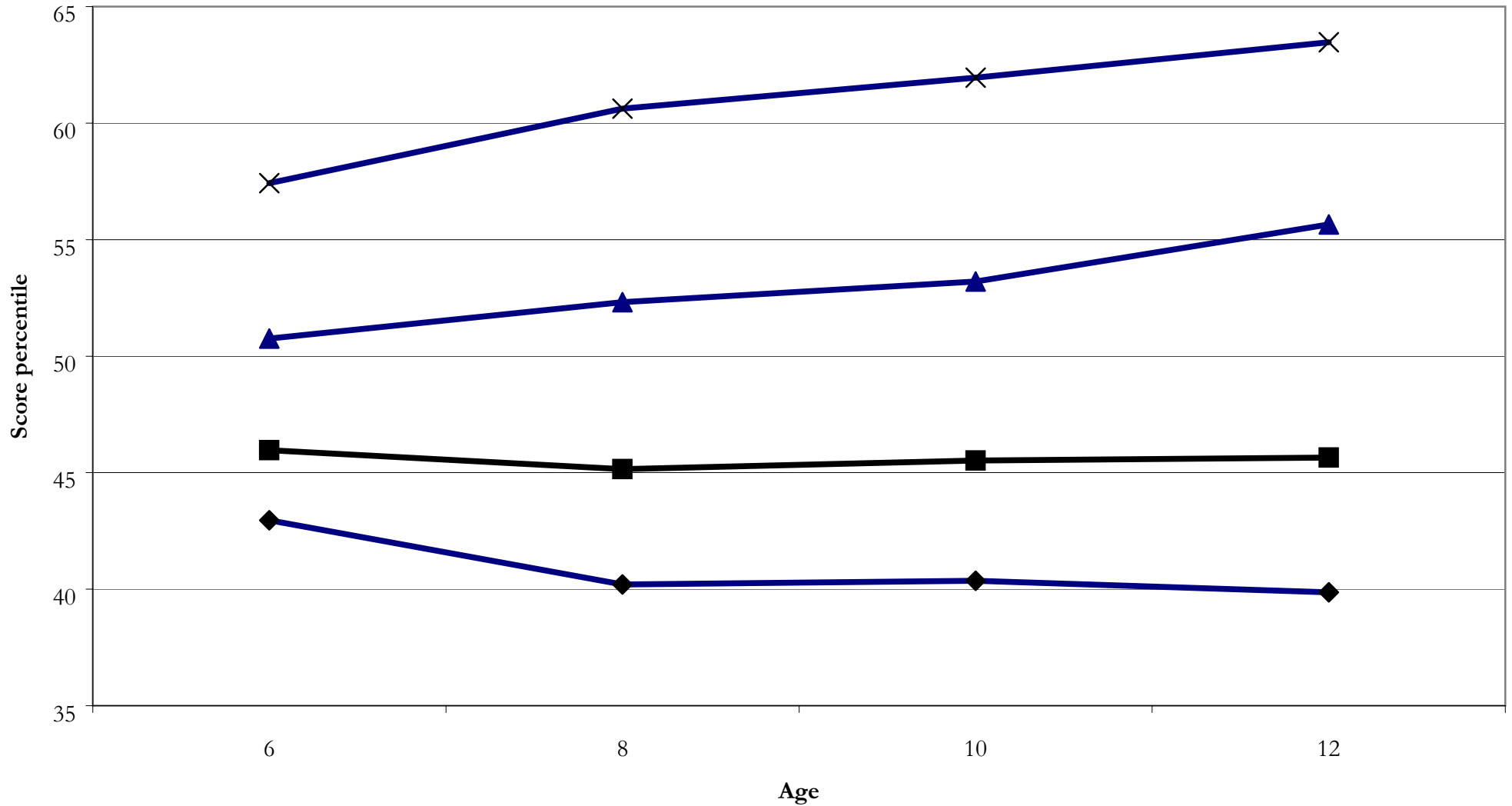
iii. By Decile of Noncognitive Factor



Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

# Figure 2A Children of NLSY

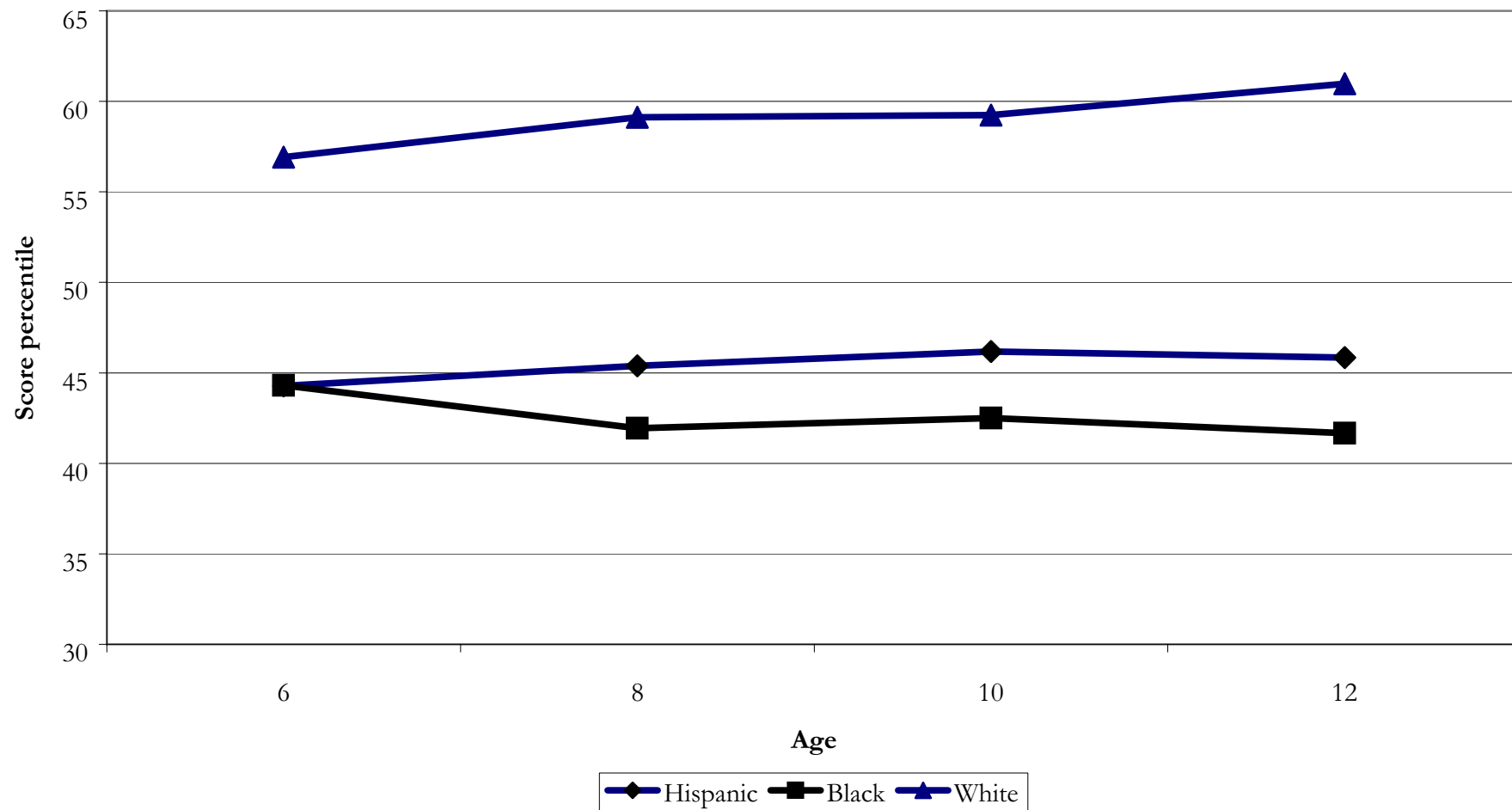
Average percentile rank on PIAT Math score, by income quartile\*



\*Income quartiles are computed from average family income between the ages of 6 and 10.

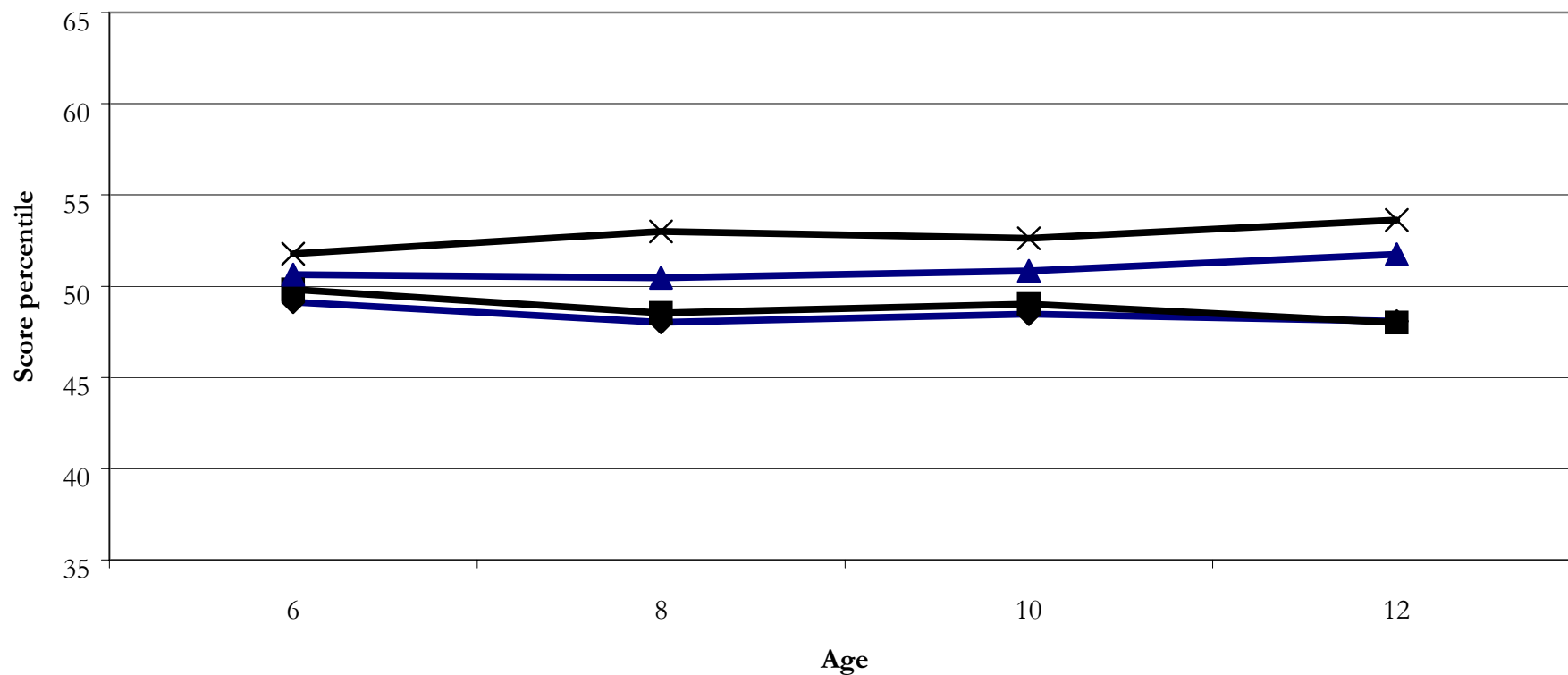
◆ Lowest income quartile   ■ Second income quartile   ▲ Third income quartile   ✕ Highest income quartile

Figure 2B  
Average percentile rank on PIAT Math score, by race



# Figure 3A Children of NLSY

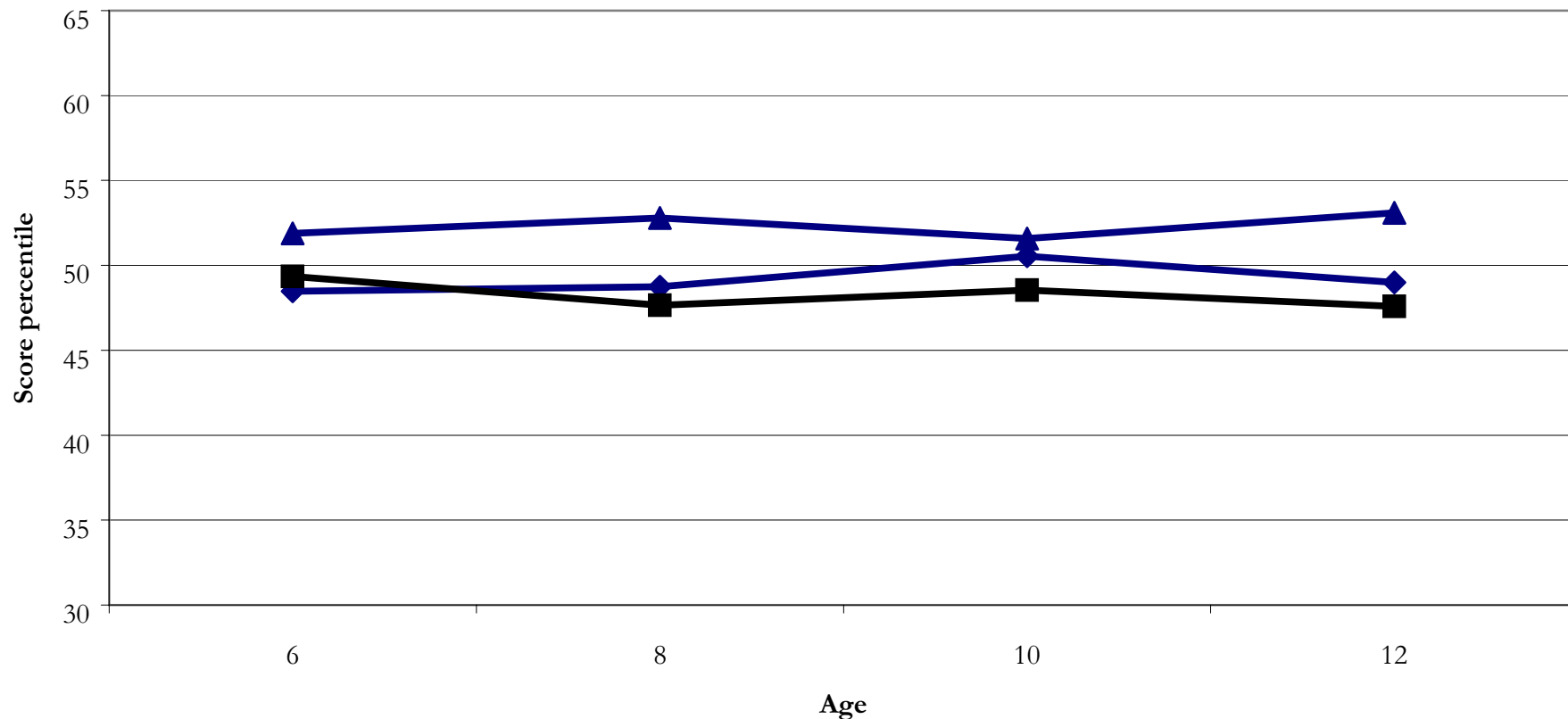
Adjusted average PIAT Math score percentiles by income quartile\*



\* Adjusted by maternal education, maternal AFQT (corrected for the effect of schooling) and broken home at each age

◆ Lowest income quartile   ■ Second income quartile   ▲ Third income quartile   ✕ Highest income quartile

Figure 3B  
Adjusted average PIAT Math score percentile by race\*



\* Adjusted by maternal education, maternal AFQT (corrected for the effect of schooling) and broken home at each age

◆ Hispanic ■ Black ▲ White

# Figure 4A Children of NLSY

Average percentile rank on anti-social behavior score, by income quartile\*

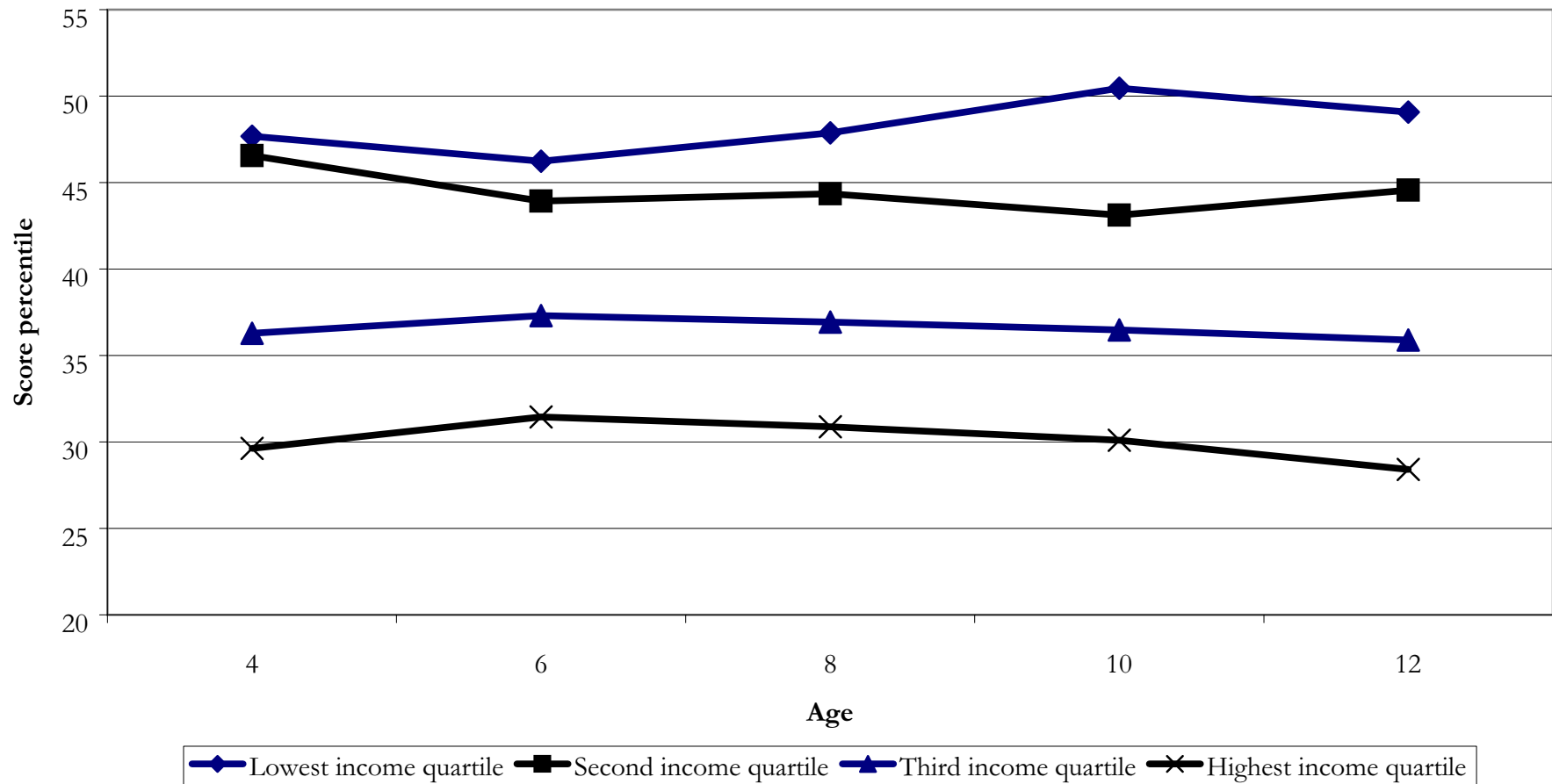
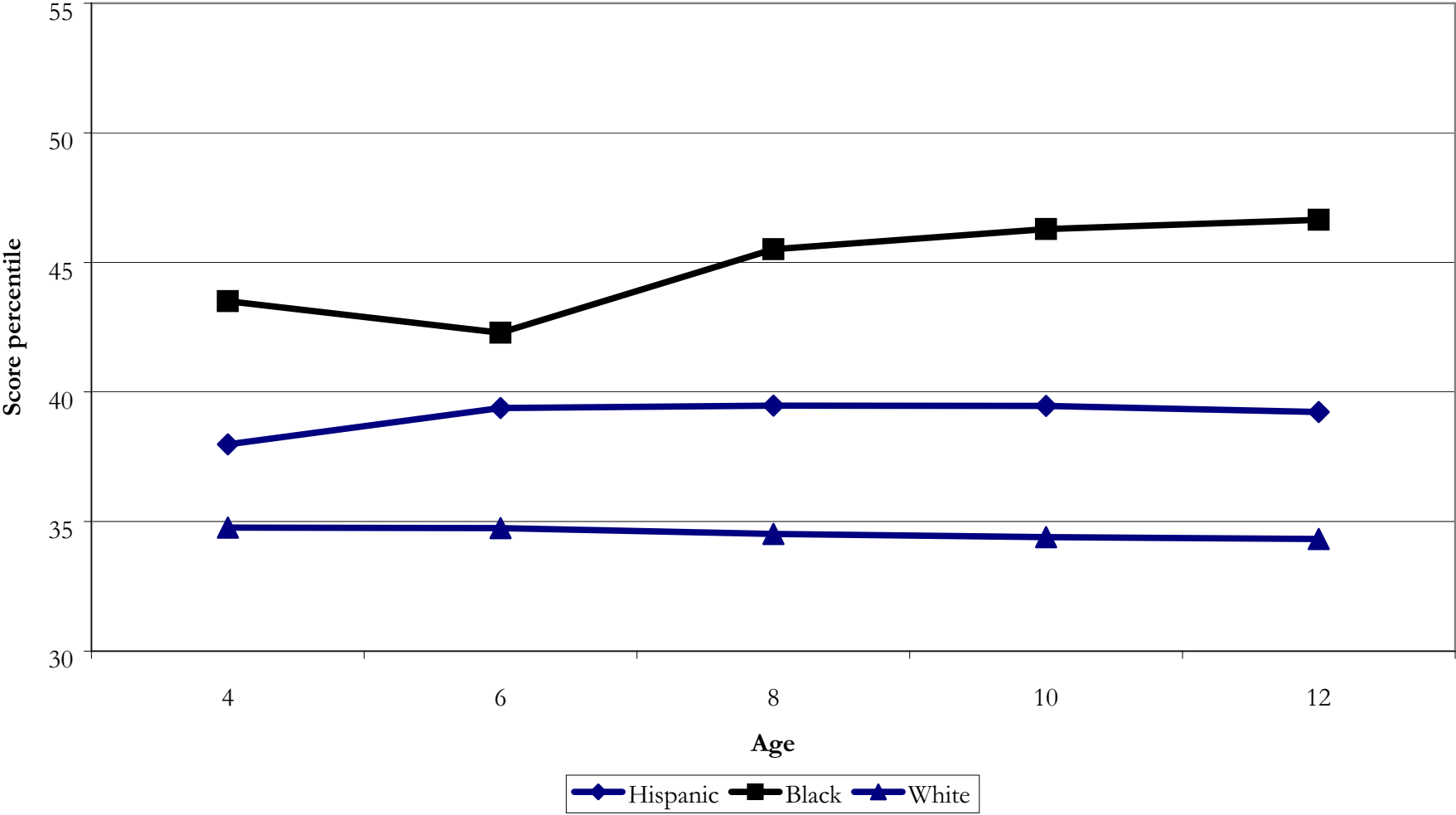
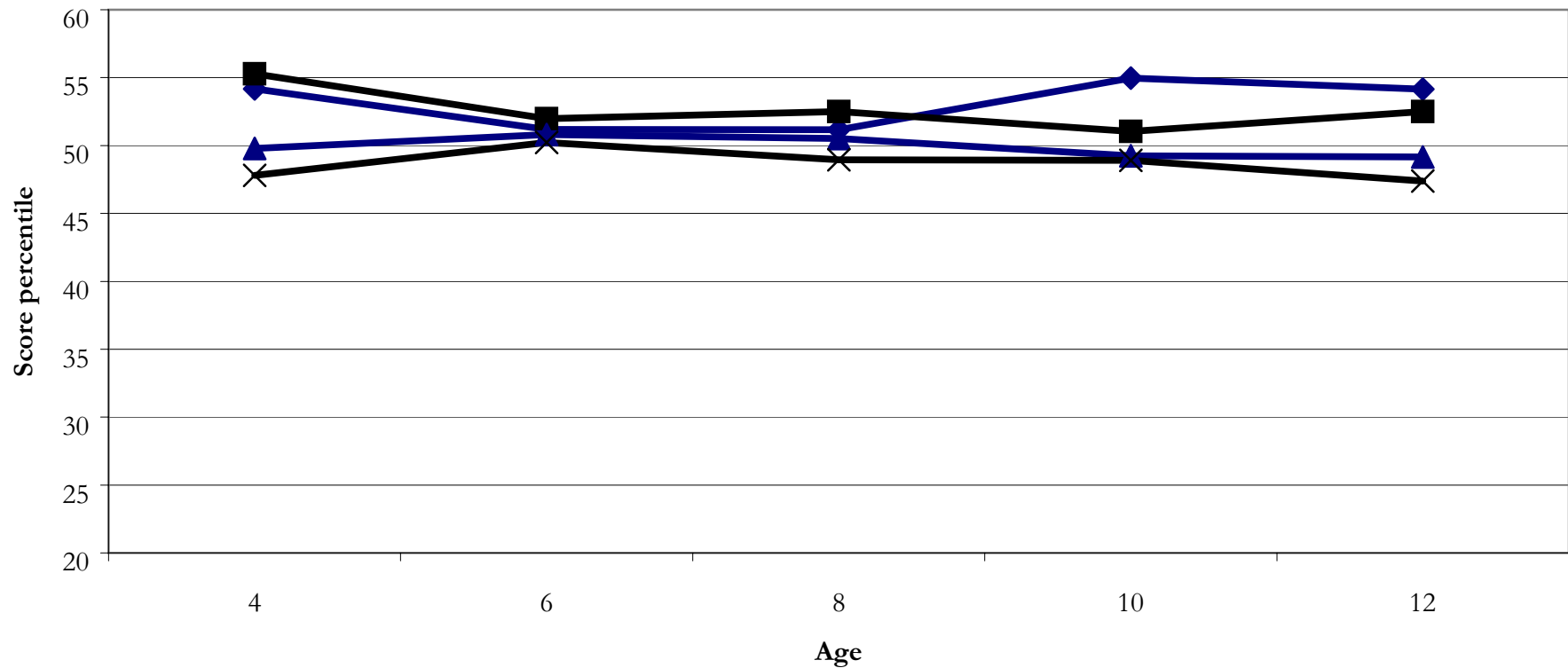


Figure 4B  
Average percentile rank on anti-social behavior score, by race



# Figure 5A Children of NLSY

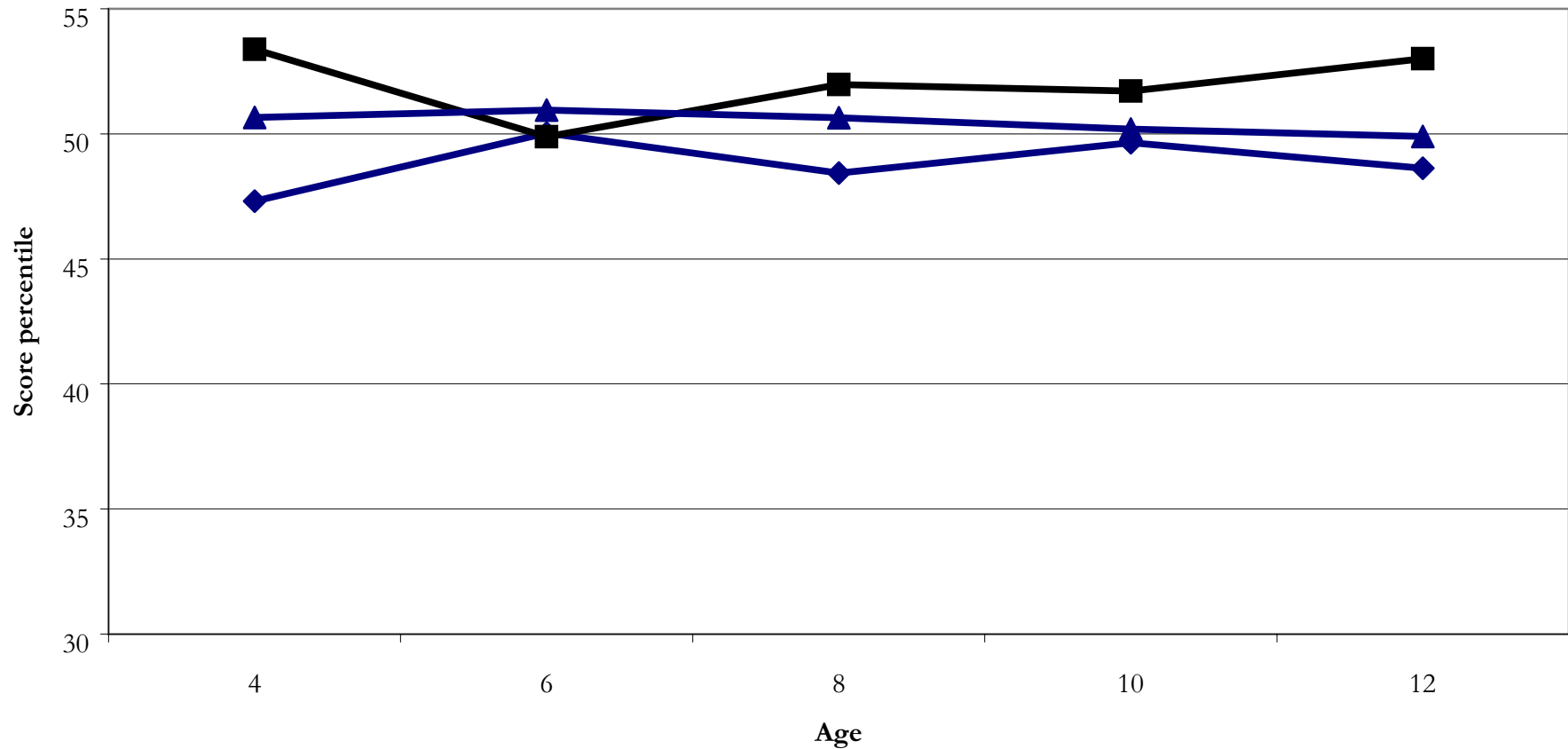
Adjusted average anti-social behavior score percentile by income quartile\*



\* Adjusted by maternal education, maternal AFQT (corrected for the effect of schooling) and broken home at each age

◆ Lowest income quartile   ■ Second income quartile   ▲ Third income quartile   ✕ Highest income quartile

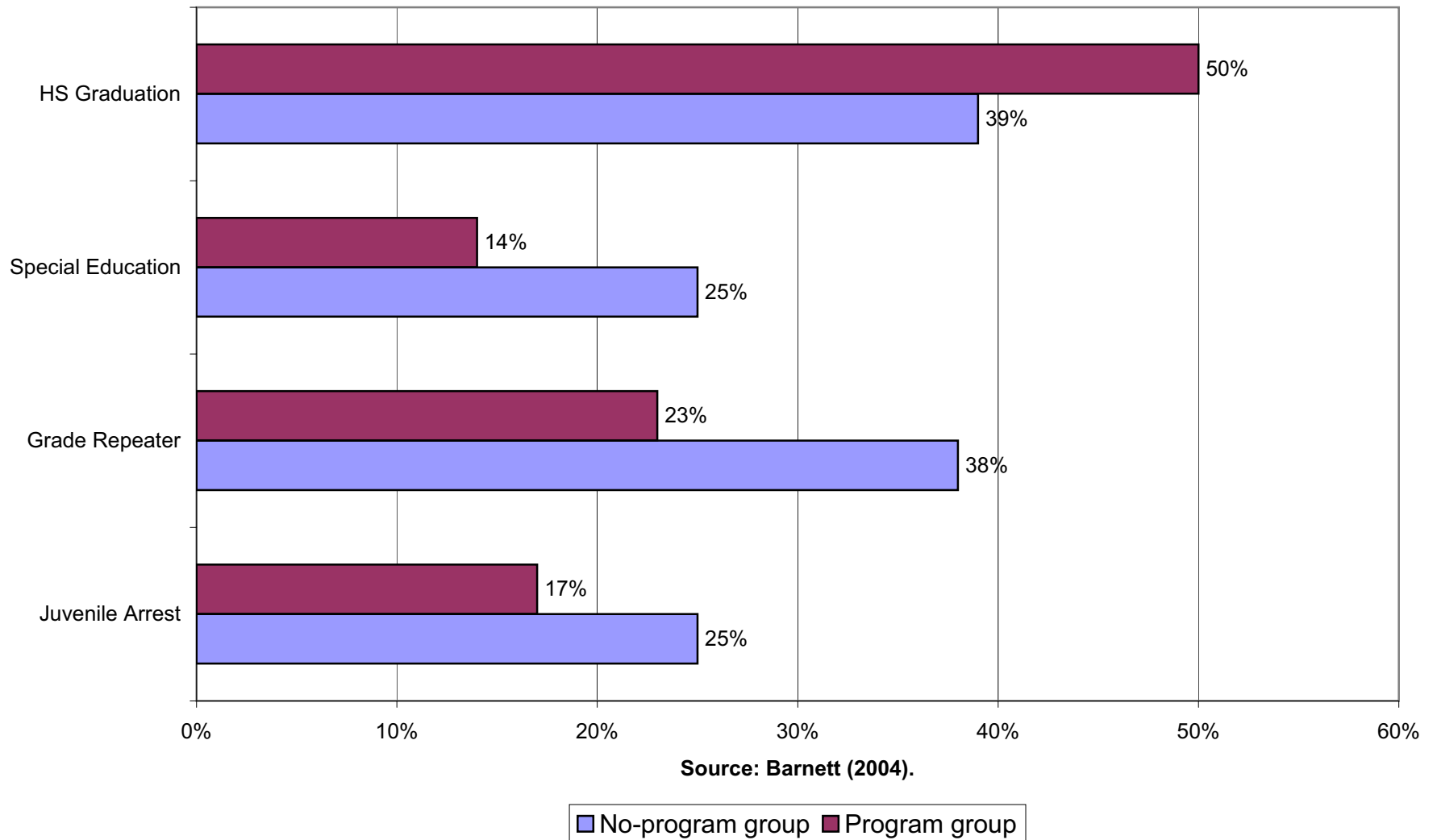
Figure 5B  
Adjusted average anti-social behavior score percentile by race\*



\* Adjusted by maternal education, maternal AFQT (corrected for the effect of schooling) and broken home at each age

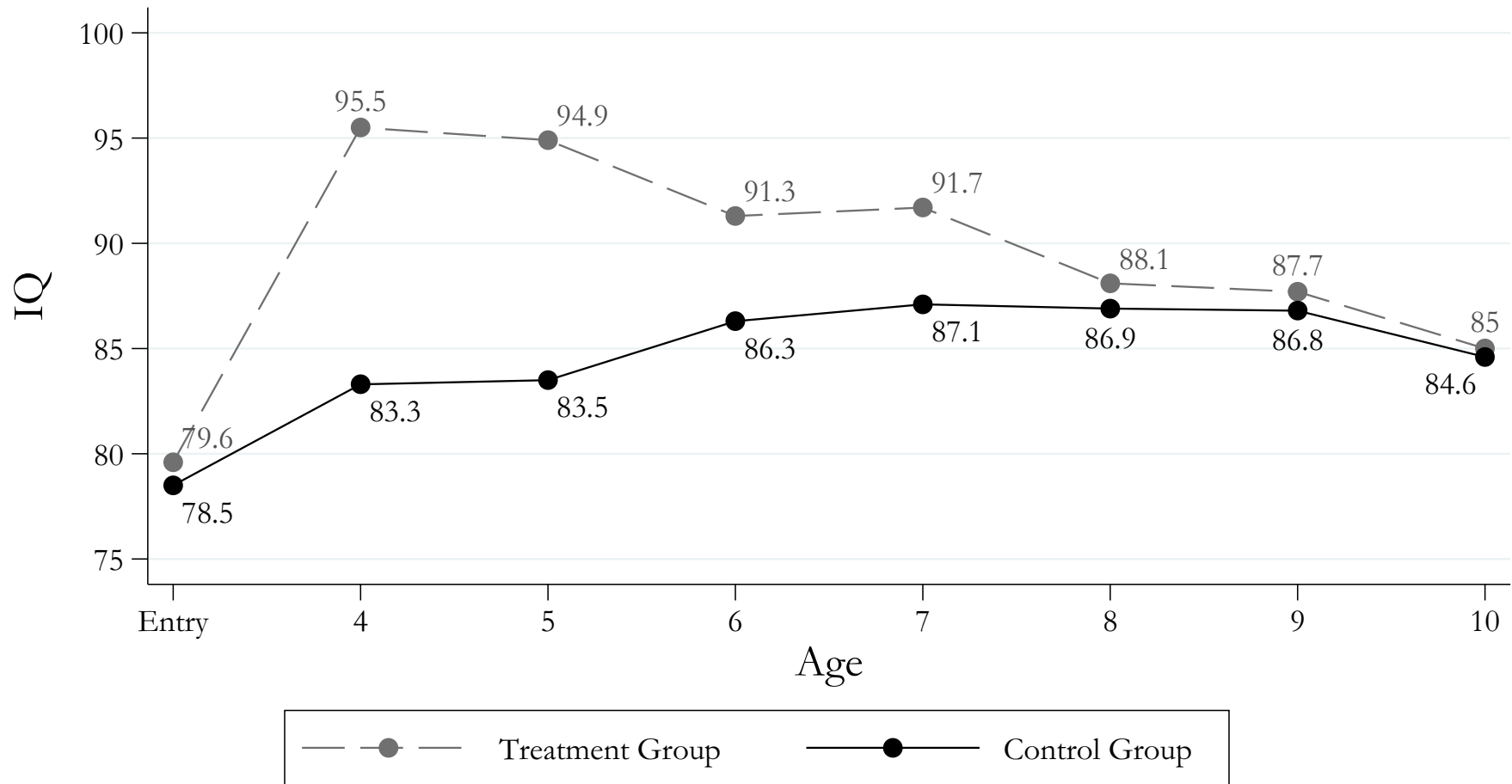
◆ Hispanic ■ Black ▲ White

**Figure 6**  
**Academic and Social Benefits at School Exit For CPC Participants**



# Figure 7A

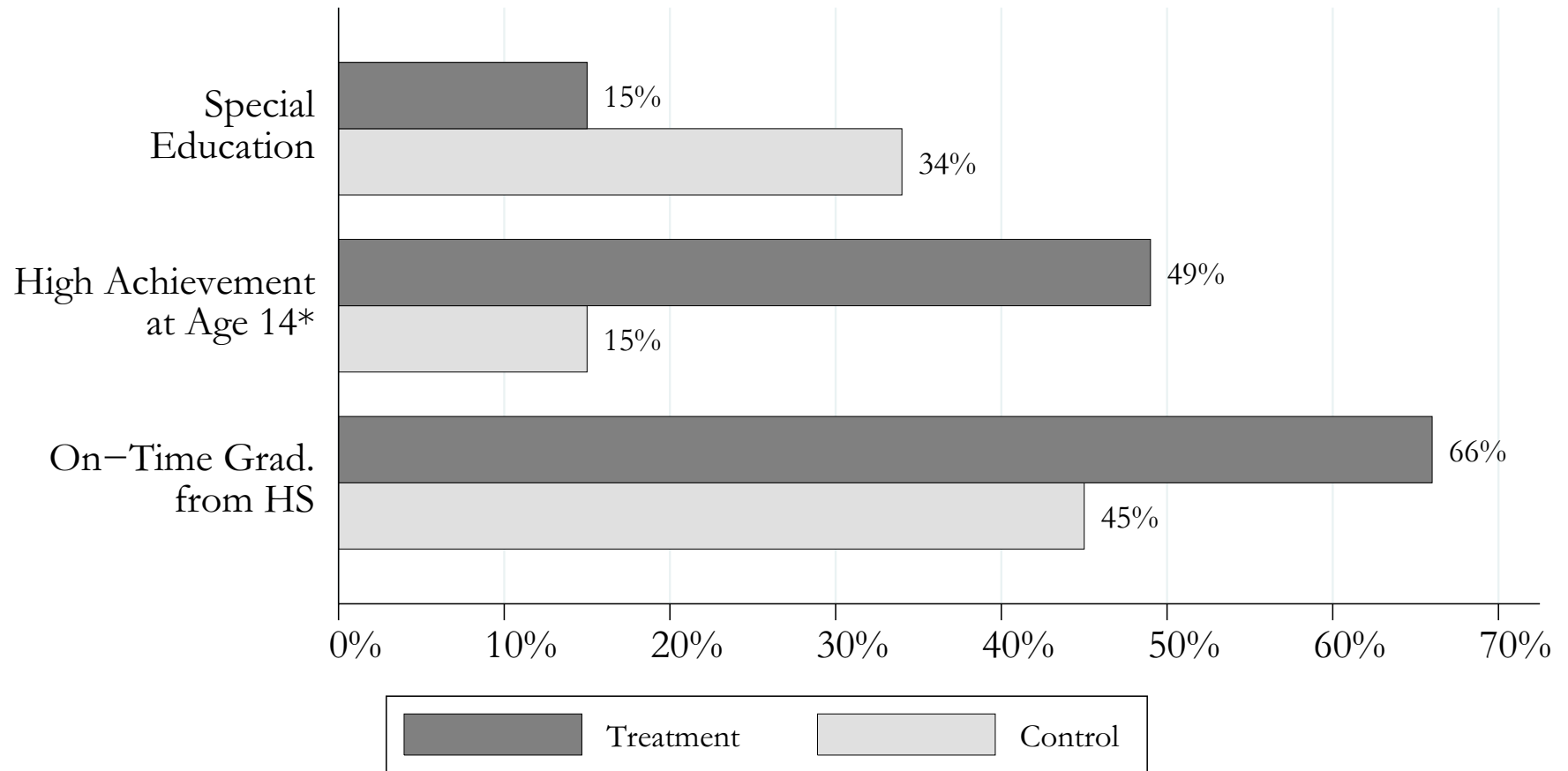
## Perry Preschool Program: IQ, by Age and Treatment Group



Source: Perry Preschool Program. IQ measured on the Stanford-Binet Intelligence Scale (Terman & Merrill, 1960). Test was administered at program entry and each of the ages indicated.

# Figure 7B

## Perry Preschool Program: Educational Effects, by Treatment Group

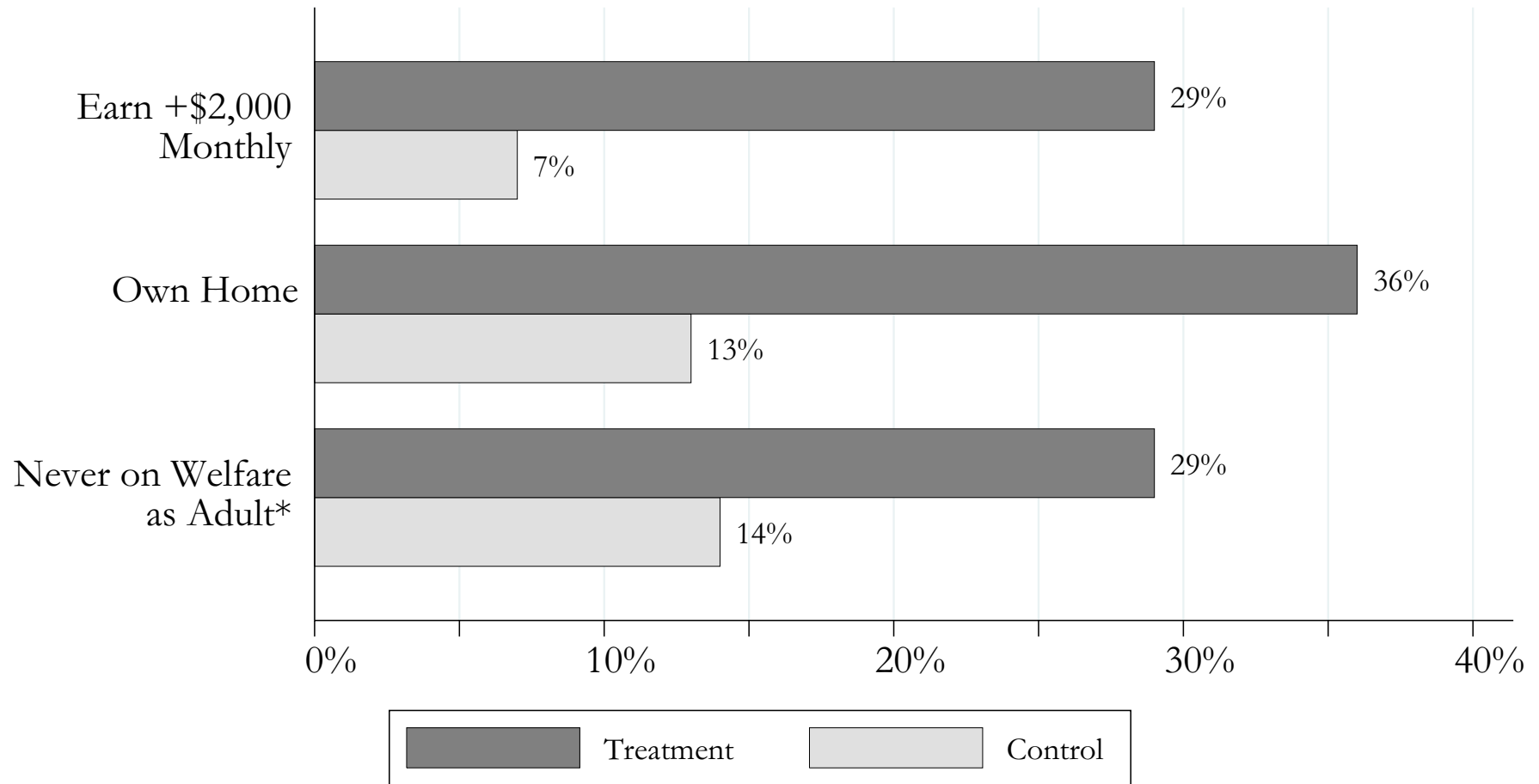


Source: Barnett (2004).

Notes: \*High achievement defined as performance at or above the lowest 10th percentile on the California Achievement Test (1970).

# Figure 7C

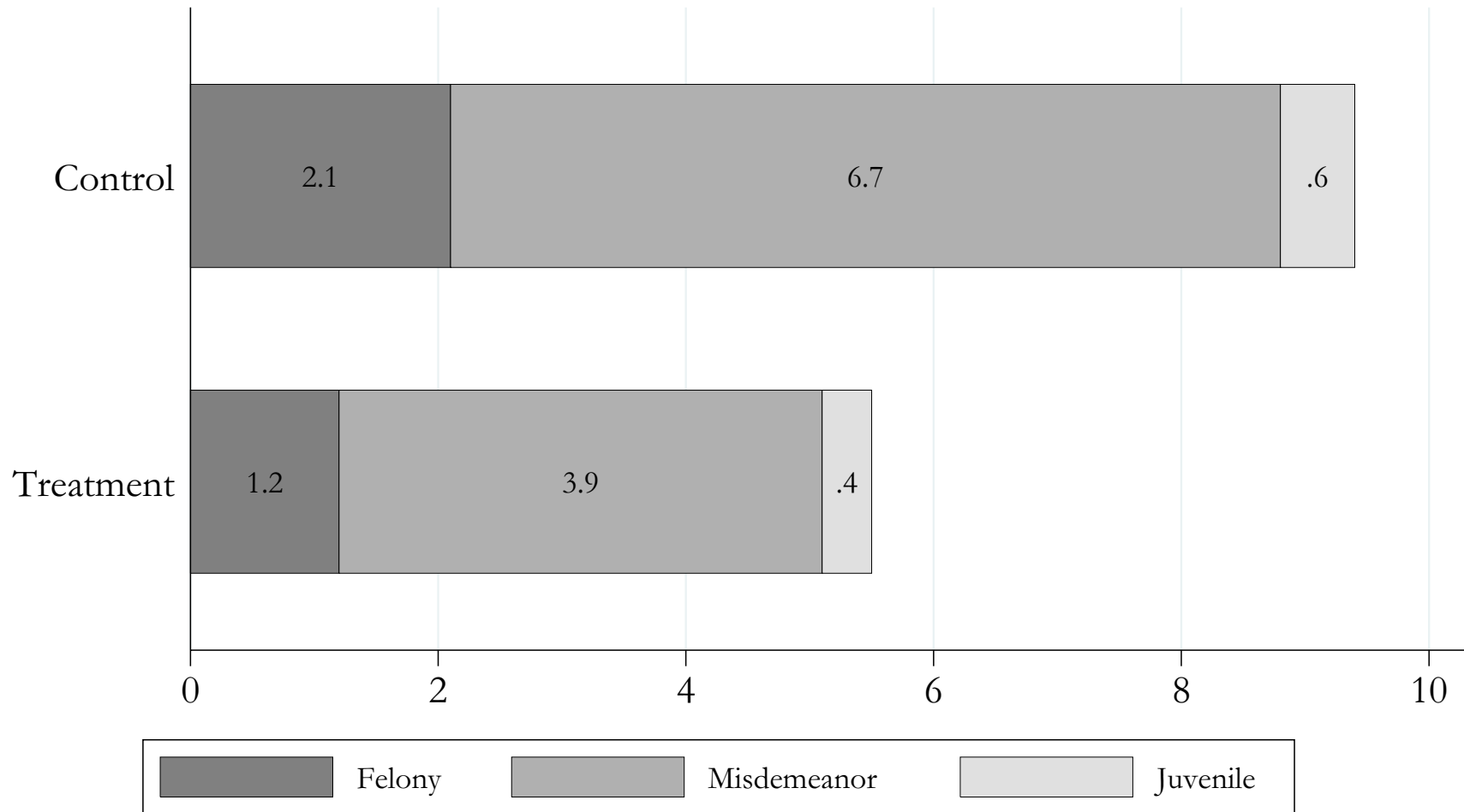
## Perry Preschool Program: Economic Effects at Age 27, by Treatment Group



Source: Barnett (2004). \*Updated through Age 40 using recent Perry Preschool Program data, derived from self-report and all available state records.

# Figure 7D

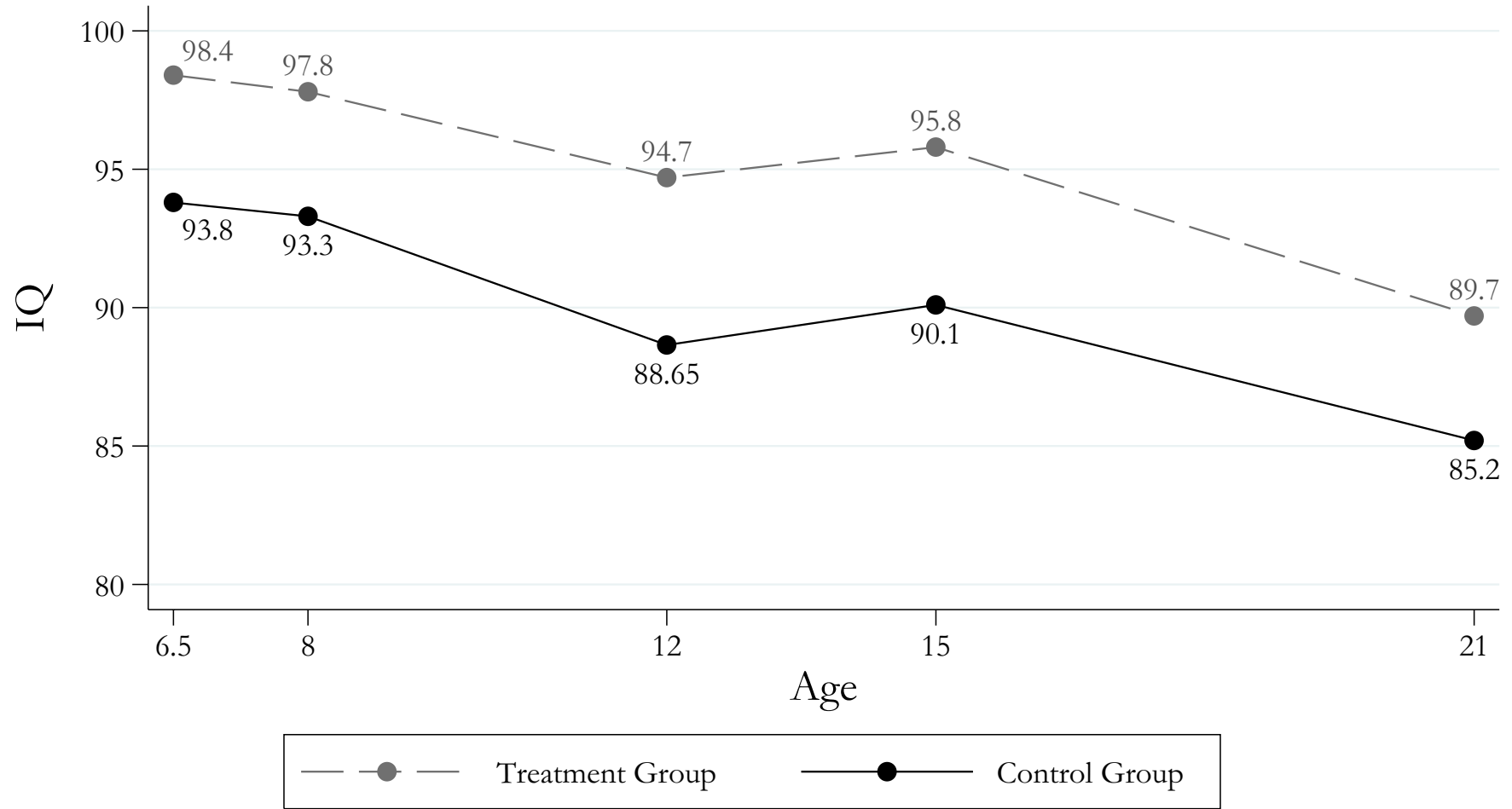
## Perry Preschool Program: Arrests per Person before Age 40, by Treatment Group



Source: Perry Preschool Program. Juvenile arrests are defined as arrests prior to age 19.

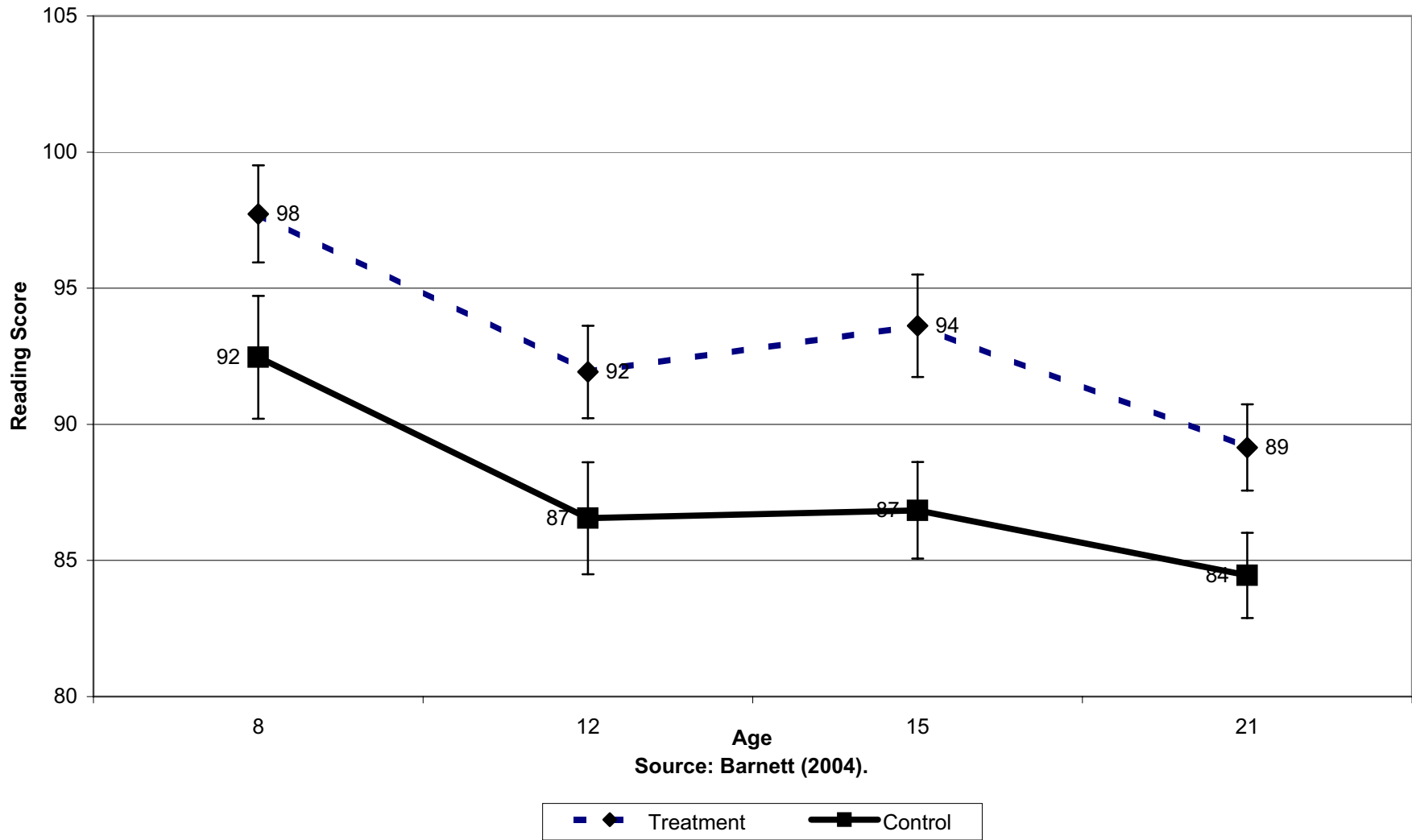
# Figure 8A

## Abecedarian Program: IQ, by Age and Treatment Group

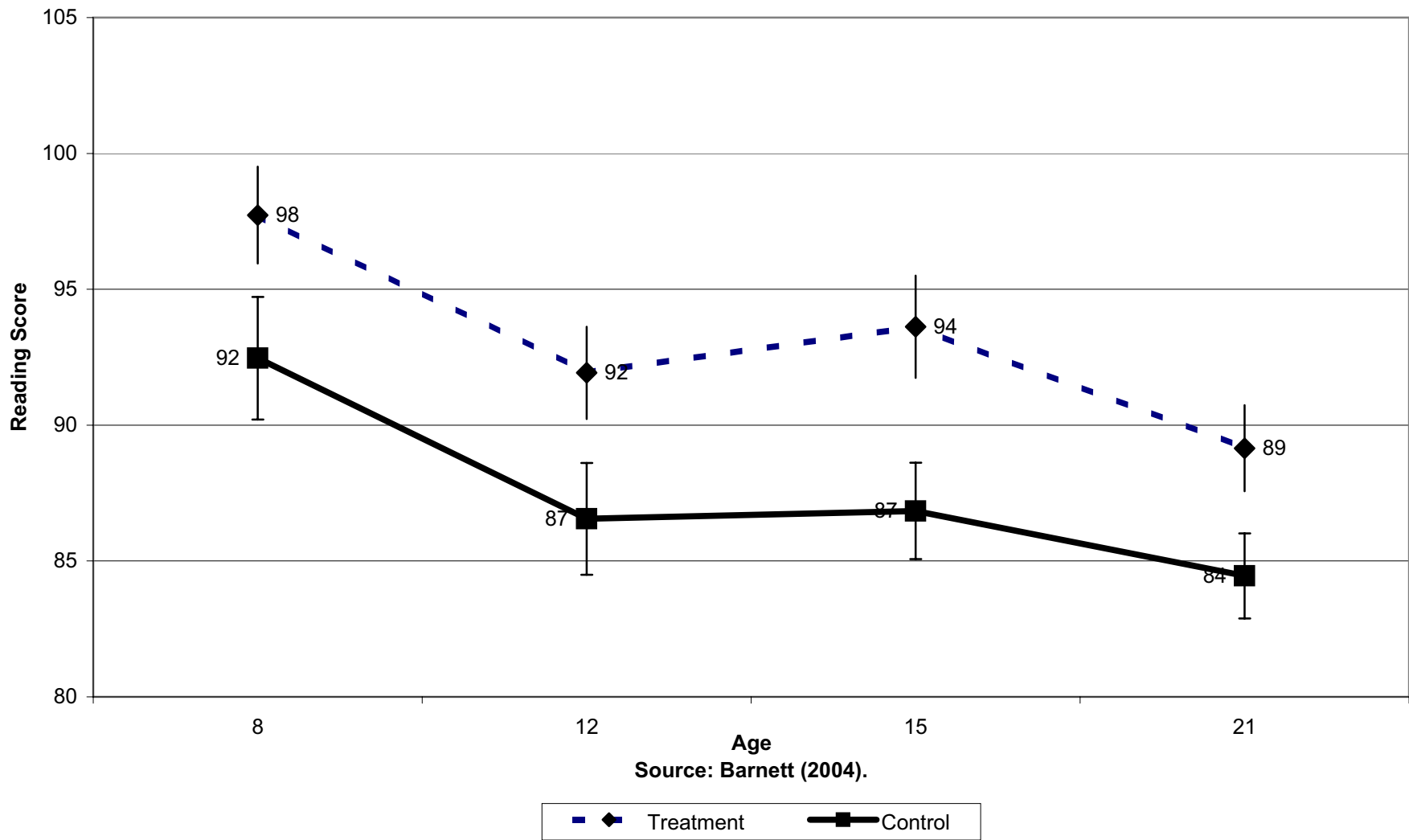


Source: Barnett (2004).

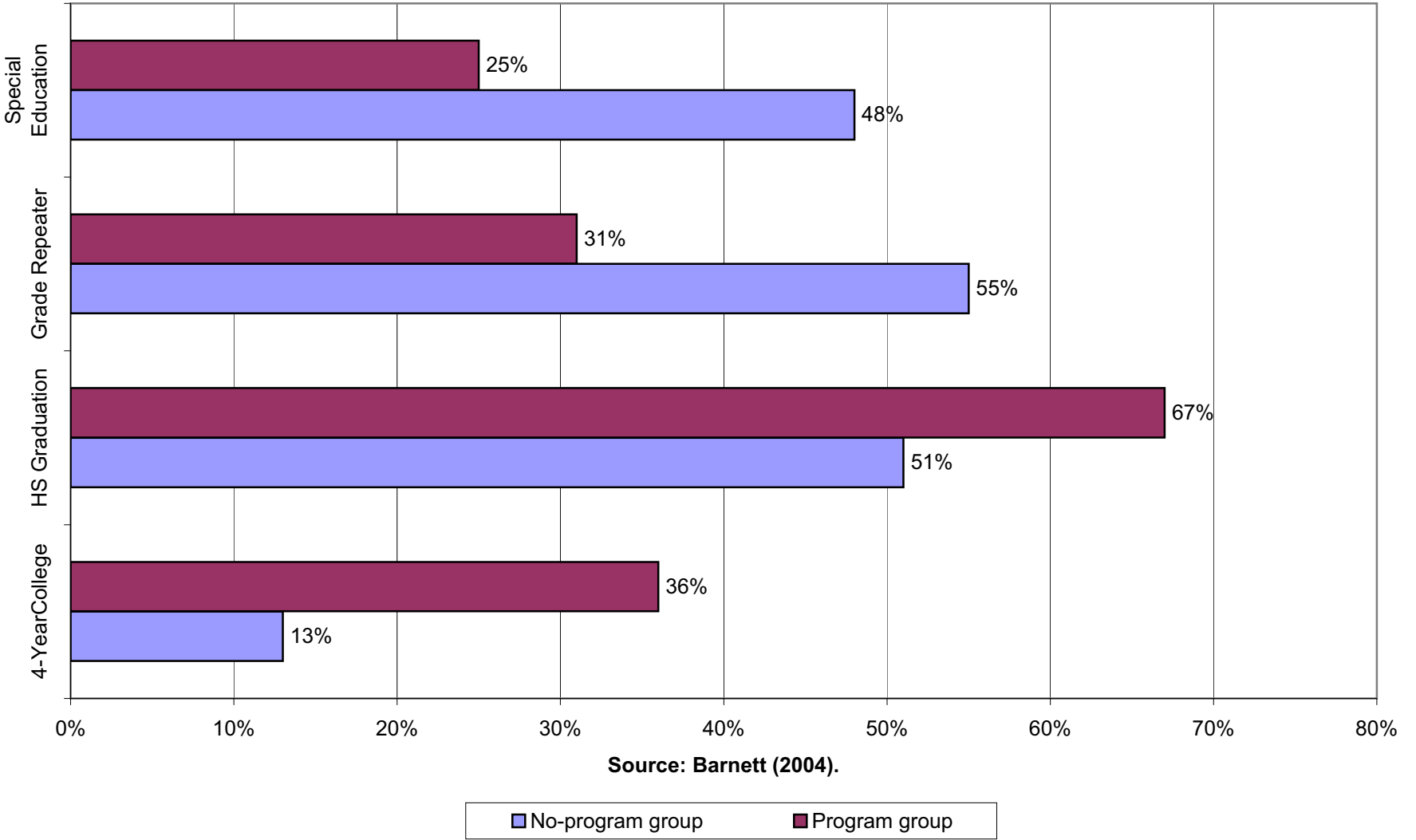
**Figure 8B**  
**Abecedarian Reading Achievement Over Time**



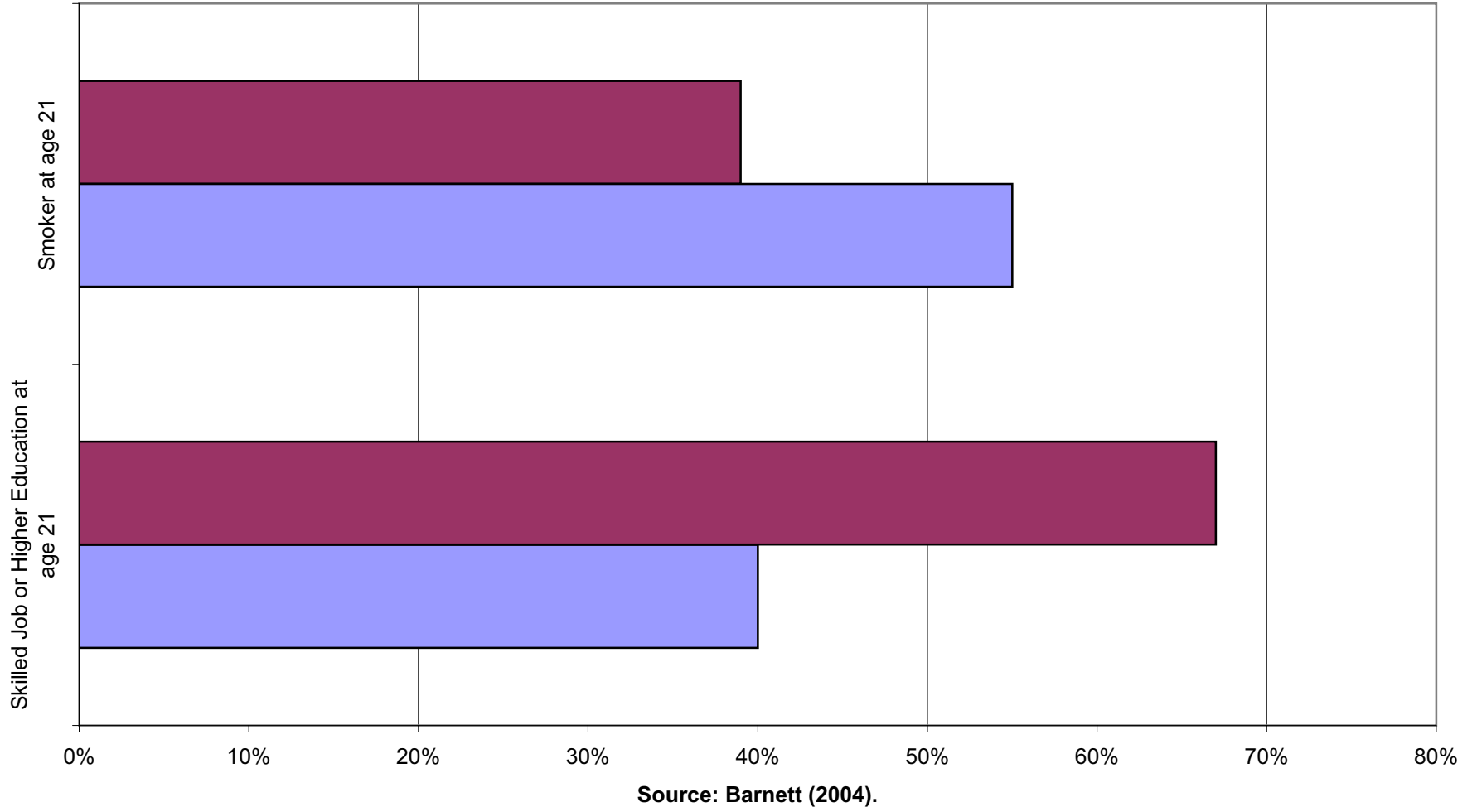
**Figure 8C**  
**Abecedarian Math Achievement Over Time**



**Figure 8D**  
**Abecedarian Academic Outcomes**



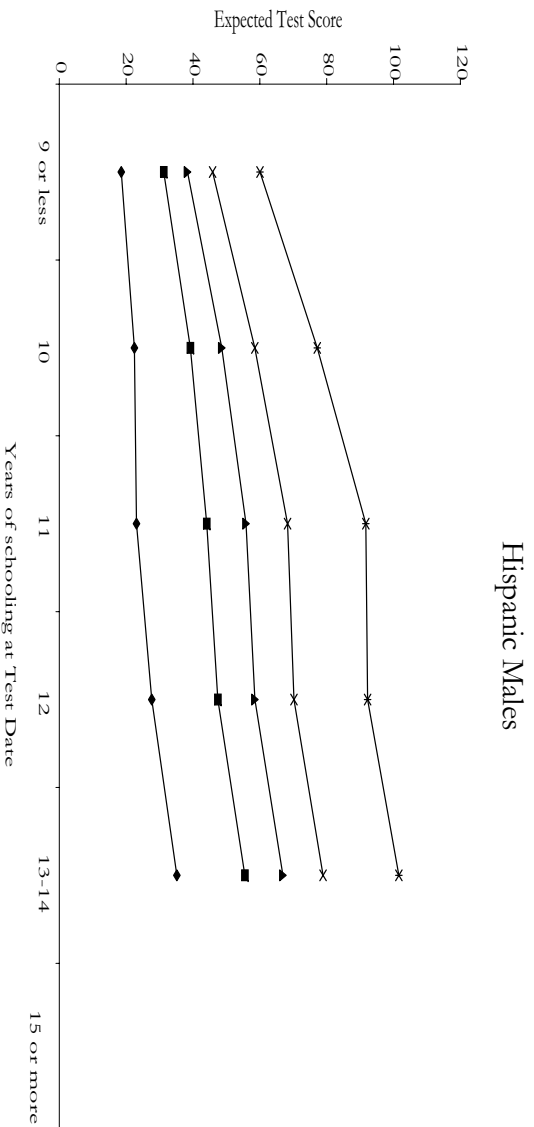
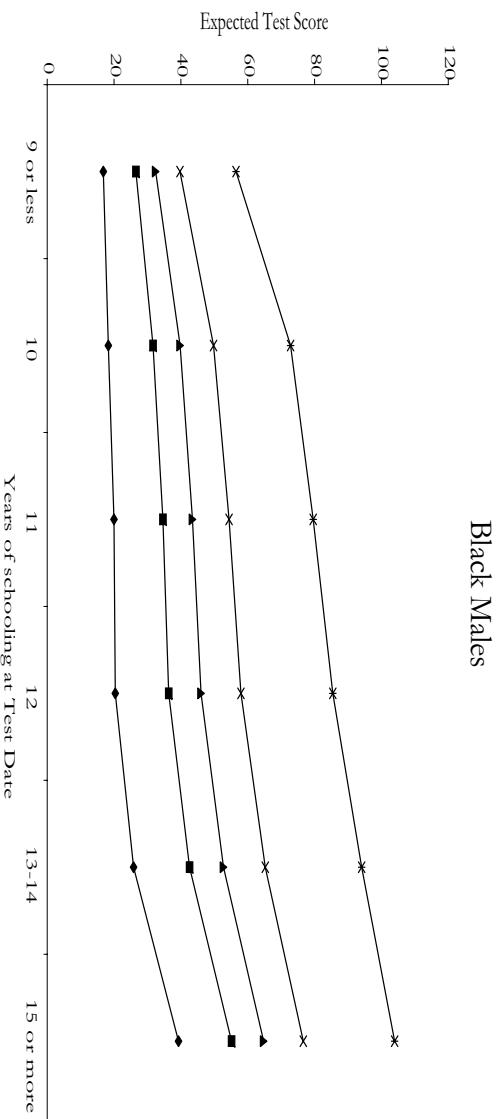
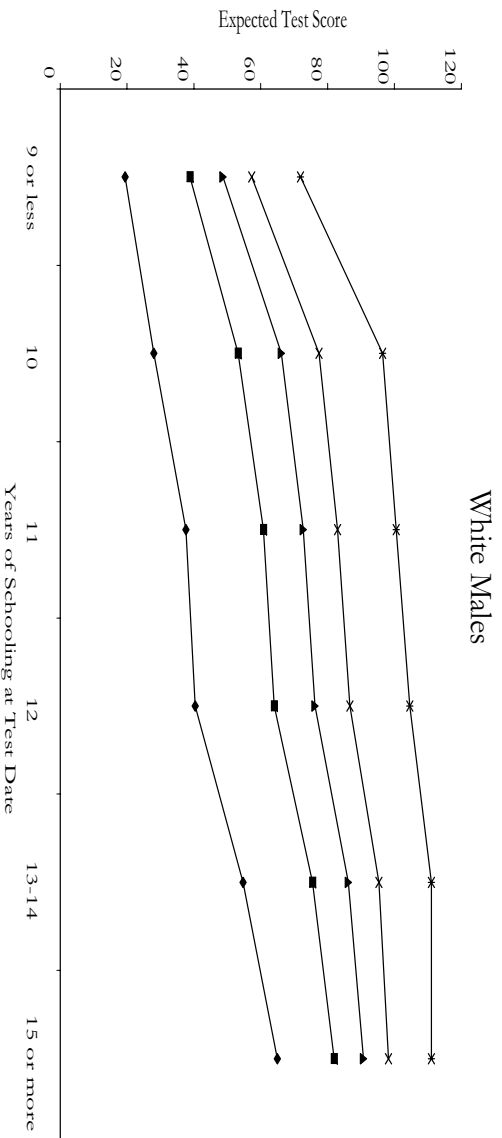
**Figure 8E**  
**Other Benefits of Abecedarian**



Source: Barnett (2004).

■ No-program group    ■ Program group

**Figure 9A**  
**Expected Value of Achievement Scores Conditional on Factor**

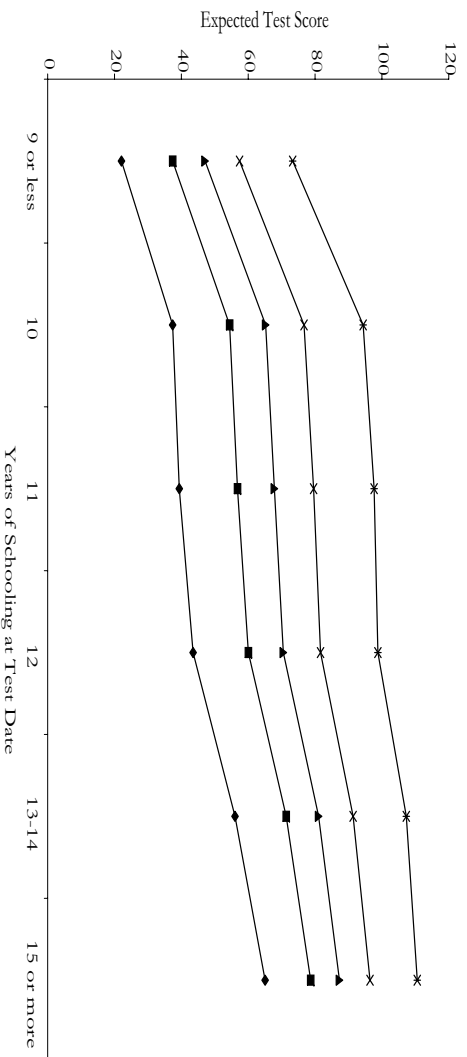


2.5th Percentile
  25th Percentile
  50th Percentile
  75th Percentile
  97.5th Percentile

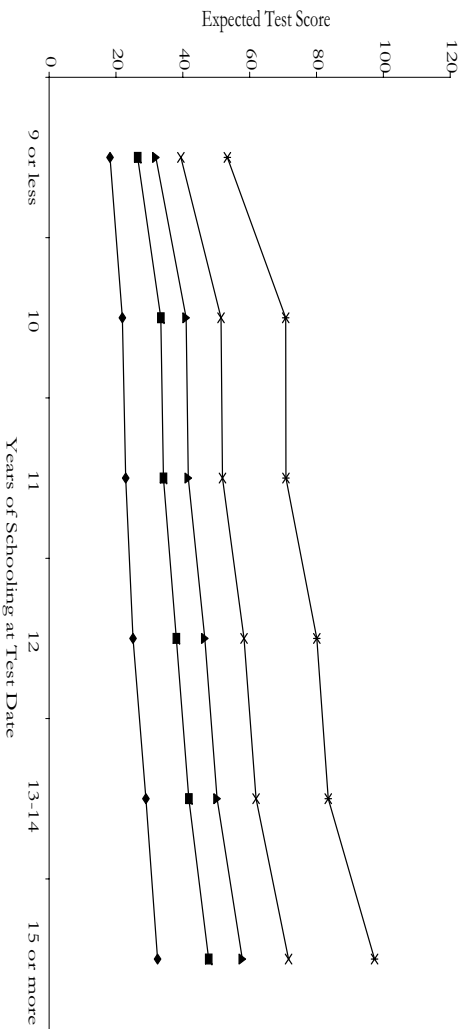
Notes: The estimates from the structural model include the following covariates: urban status, broken home and southern residence at age 14, number of siblings and family income in 1979, mother's and father's education, and age at December 1980. The measure of achievement is the Armed Forces Qualifying Test (AFQT).

Source: Heckman, Laremas and Urzua (2004)

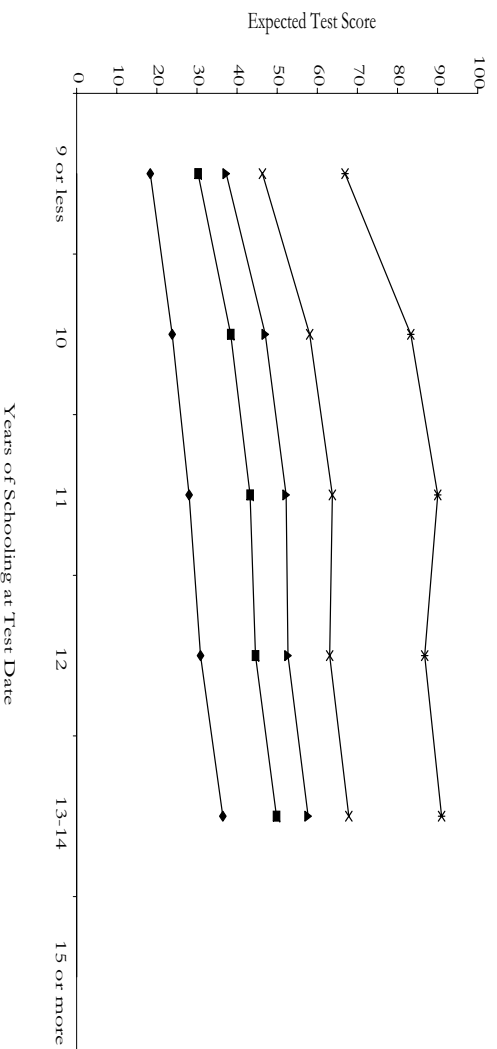
**Figure 9B**  
 Expected Value of Achievement Scores Conditional on Factor  
 White Females



**Black Females**



**Hispanic Females**

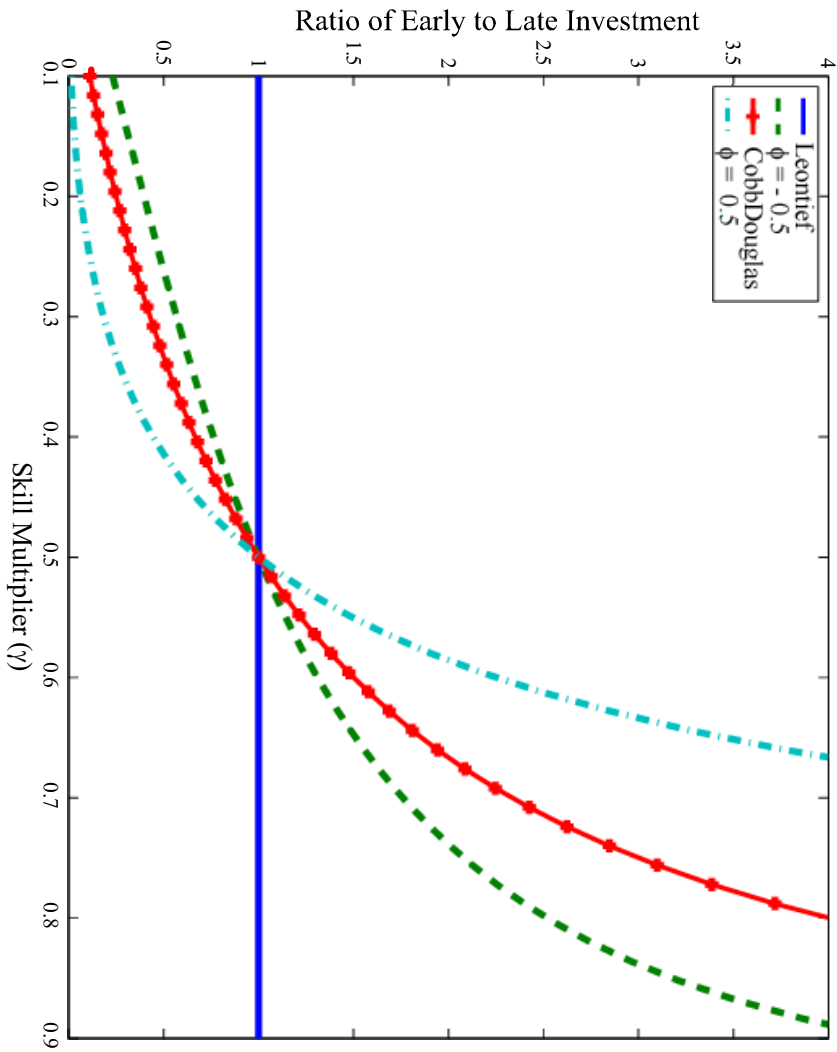


◆ 2.5th Percentile    ■ 25th Percentile    ▲ 50th Percentile    ✕ 75th Percentile    \* 97.5th Percentile

Notes: The estimates from the structural model include the following covariates: urban status, broken home and southern residence at age 14, number of siblings and family income in 1979, mother's and father's education, and age at December 1980. The measure of achievement is the Armed Forces Qualifying Test (AFQT).

Figure 10

The Ratio of Early to Late Investment in Human Capital as a Function of the Skill Multiplier for Different Values of Complementarity



This figure shows the optimal ratio of early to late investments,  $\frac{I_1}{I_2}$ , as a function of the skill multiplier parameter  $\gamma$ , for different values of the complementarity parameter  $\phi$ , assuming that the interest rate  $r$  is zero. The optimal ratio  $\frac{I_1}{I_2}$  is the solution of the parental problem of maximizing the present value of the child's wealth through investments in human capital,  $h$ , and transfers of risk-free bonds,  $b$ . In order to do that, parents have to decide how to allocate a total of  $M$  dollars into early and late investments in human capital,  $I_1$  and  $I_2$ , respectively, and risk-free bonds. Let  $q$  denote the present value as of period “3” of the future prices of one efficiency unit of human capital:  $q = \sum_{t=3}^T \frac{w_t}{(1+r)^{t-3}}$ . The parents solve

$$\max \left( \frac{1}{1+r} \right)^2 [qh + b]$$

subject to the budget constraint

$$I_1 + \frac{I_2}{(1+r)} + \frac{b}{(1+r)^2} = M$$

and the technology of skill formation:

$$h = \left[ \gamma I_1^\phi + (1-\gamma) I_2^\phi \right]^{\frac{1}{\phi}}$$

for  $0 < \rho < 1$ ,  $0 \leq \gamma \leq 1$ , and  $\phi \leq 1$ . From the first-order conditions it follows that  $\frac{I_1}{I_2} = \left[ \frac{\gamma}{(1-\gamma)(1+r)} \right]^{\frac{1}{1-\phi}}$ . This ratio is plotted in this figure when  $\phi \rightarrow -\infty$  (Leontief),  $\phi = -0.5$ ,  $\phi = 0$  (Cobb-Douglas) and  $\phi = 0.5$  and for values of the skill multiplier  $\gamma$  between 0.1 and 0.9.

Figure 11  
Out-of-Sample Prediction Using the Perry Preschool Treatment Group  
Actual Perry Data Versus Linear and Nonlinear Model Predictions

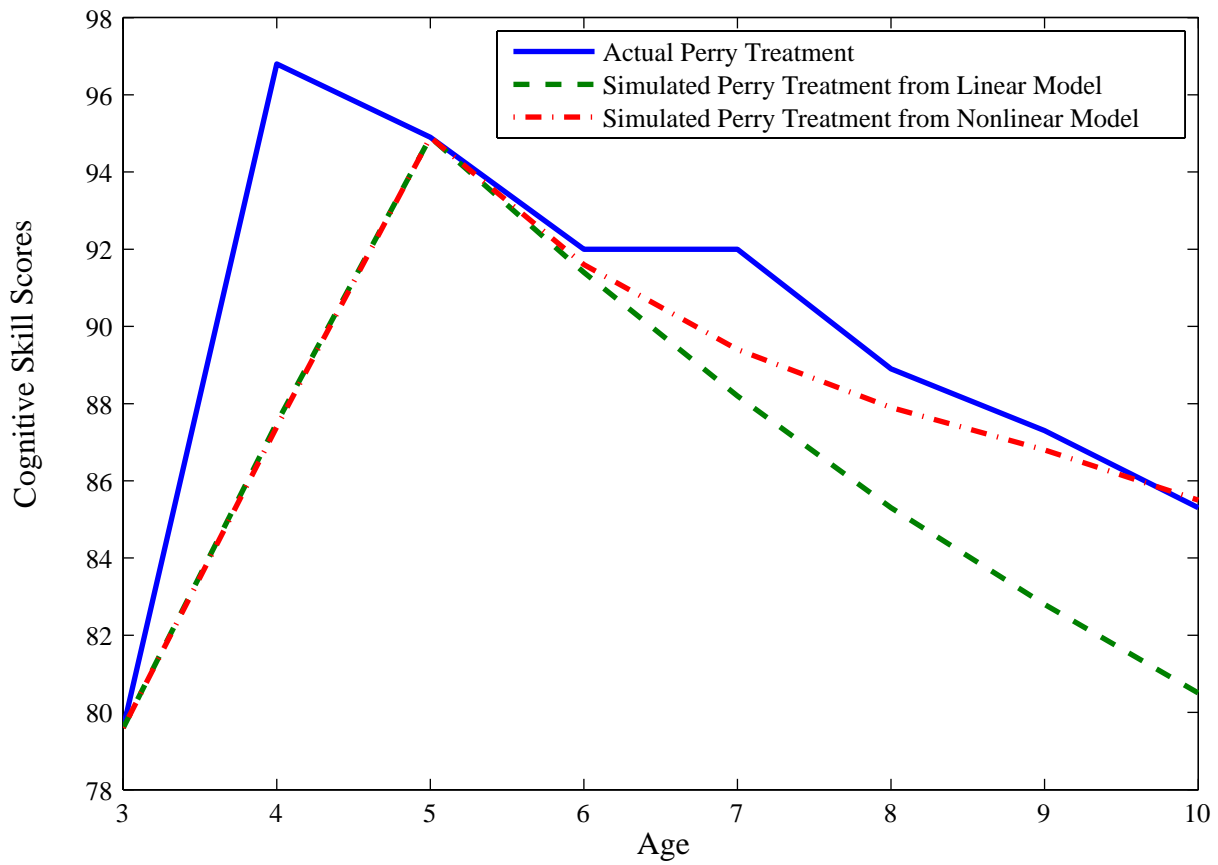
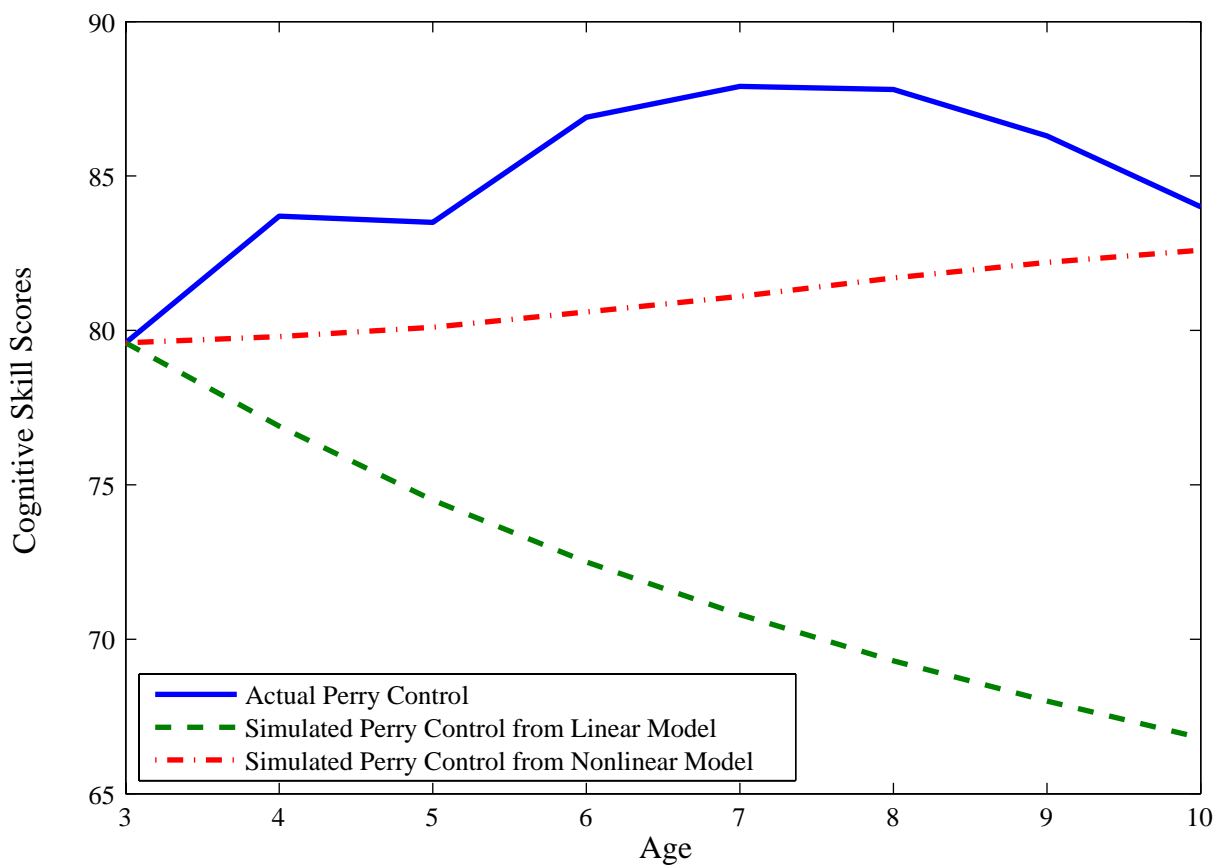
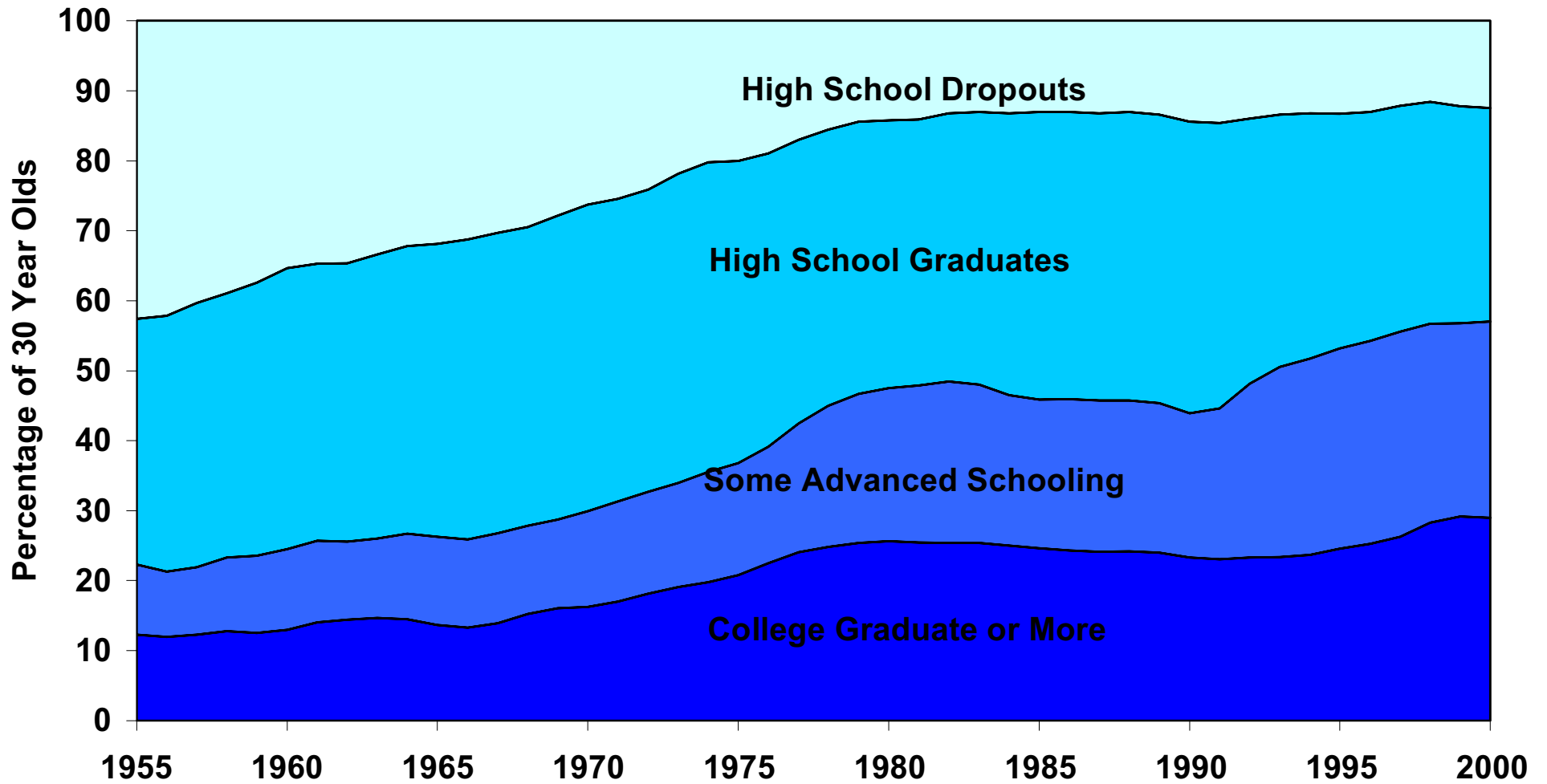


Figure 12  
Out-of-Sample Prediction Using the Perry Preschool Control Group  
Actual Perry Data Versus Linear and Nonlinear Model Predictions



**Figure 13**

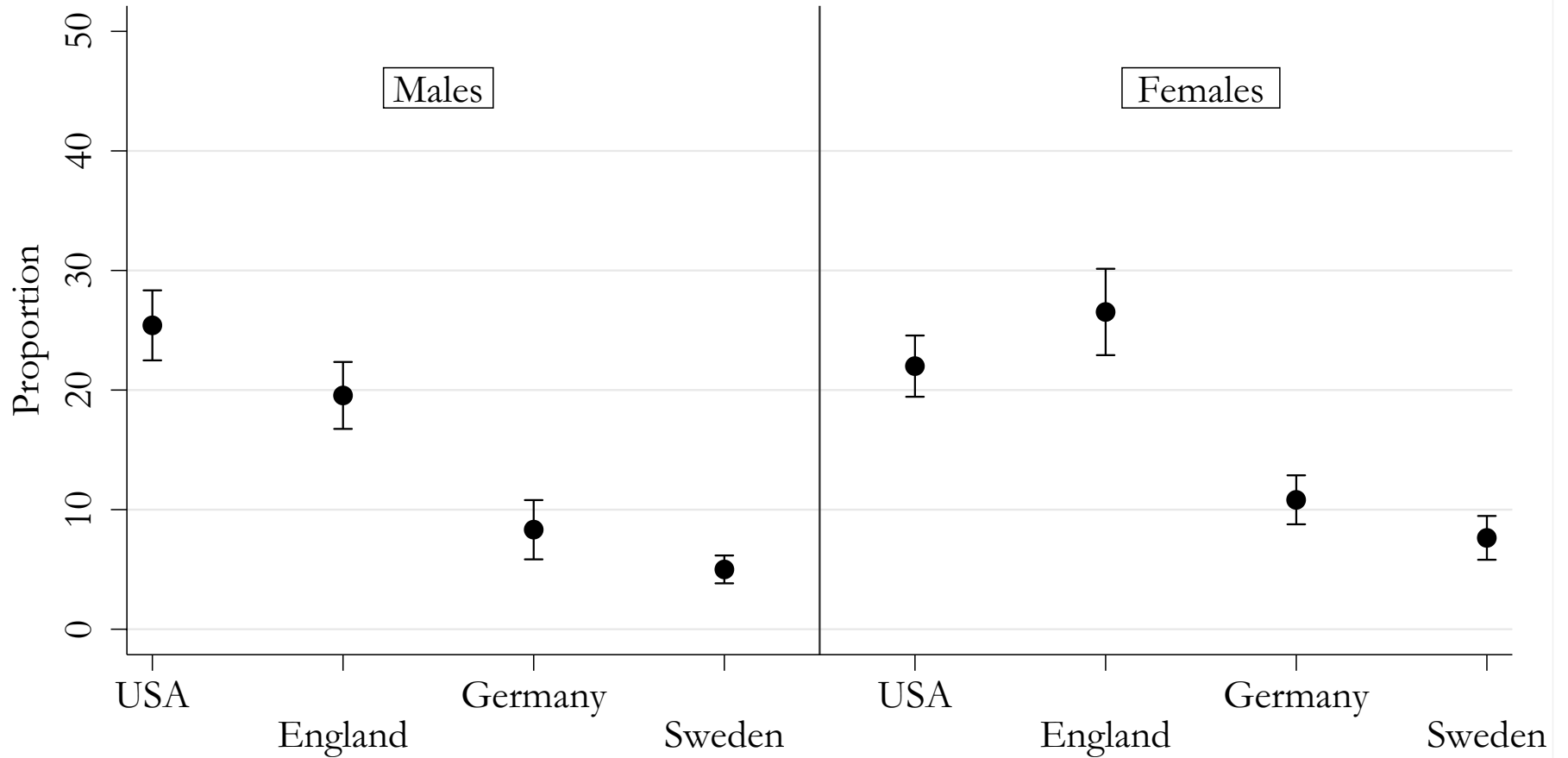
**Percent Distribution of Education Among 30 Year Olds By Year**



Source: Annual March CPS Data. Three Year Centered Moving Averages (Ellwood, 2001)

# Figure 14

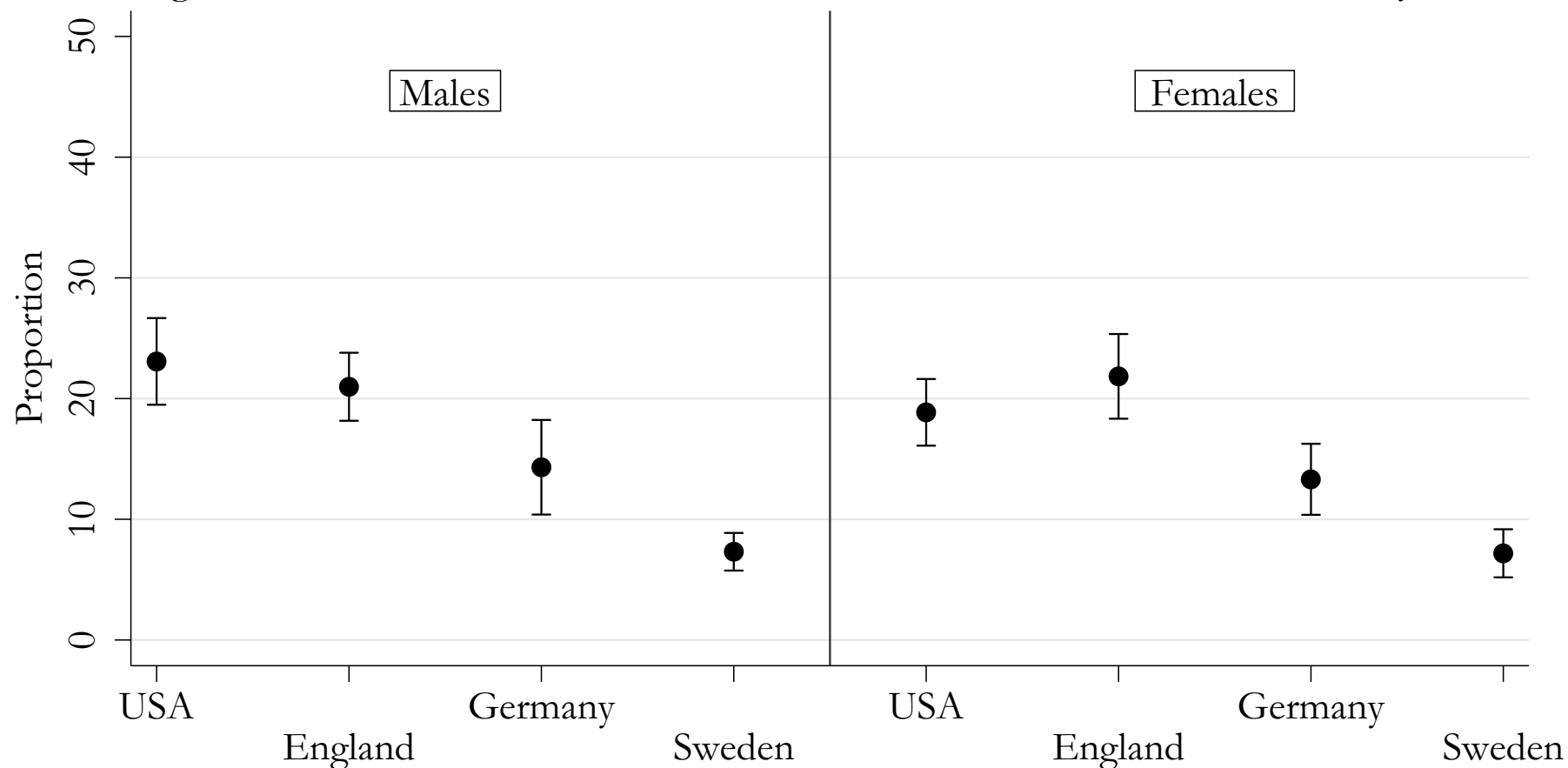
## Percentage of Each Gender Who Perform at Level 1 on the IALS Document Literacy Scale



Note: The scale scores were grouped into five levels of increasing difficulty, with Level 1 representing functional illiteracy. Levels 4 and 5 were combined. The sample is restricted to adults who are between 16–65 years of age at the time of the survey (1994 for the US and Germany, 1996 for the UK, and 1994–1995 for Sweden). Standard errors are calculated using the methodology described in IALS (2002).

# Figure 15

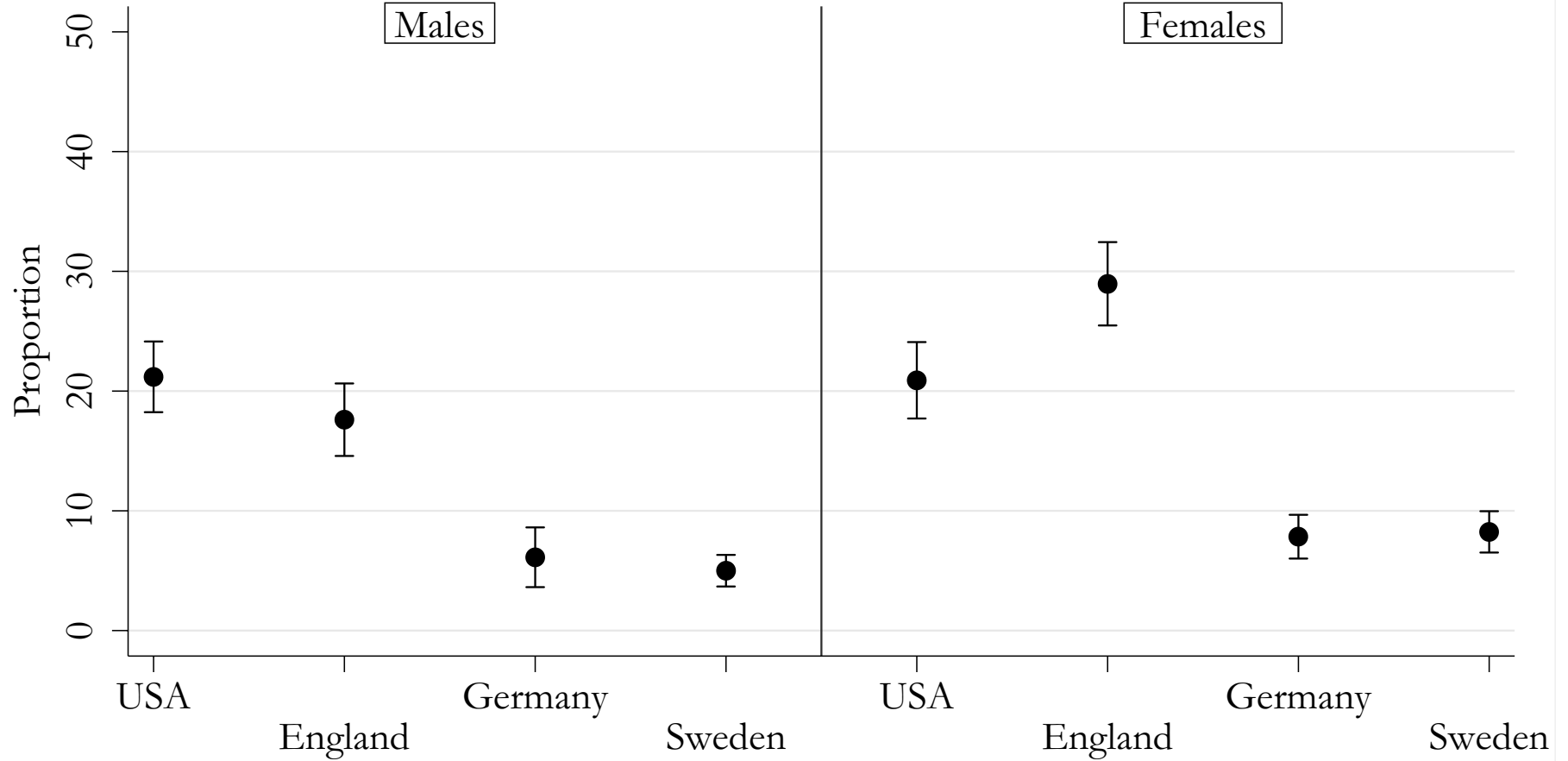
## Percentage of Each Gender Who Perform at Level 1 on the IALS Prose Literacy Scale



Note: The scale scores were grouped into five levels of increasing difficulty, with Level 1 representing functional illiteracy. Levels 4 and 5 were combined. The sample is restricted to adults who are between 16–65 years of age at the time of the survey (1994 for the US and Germany, 1996 for the UK, and 1994–1995 for Sweden). Standard errors are calculated using the methodology described in IALS (2002).

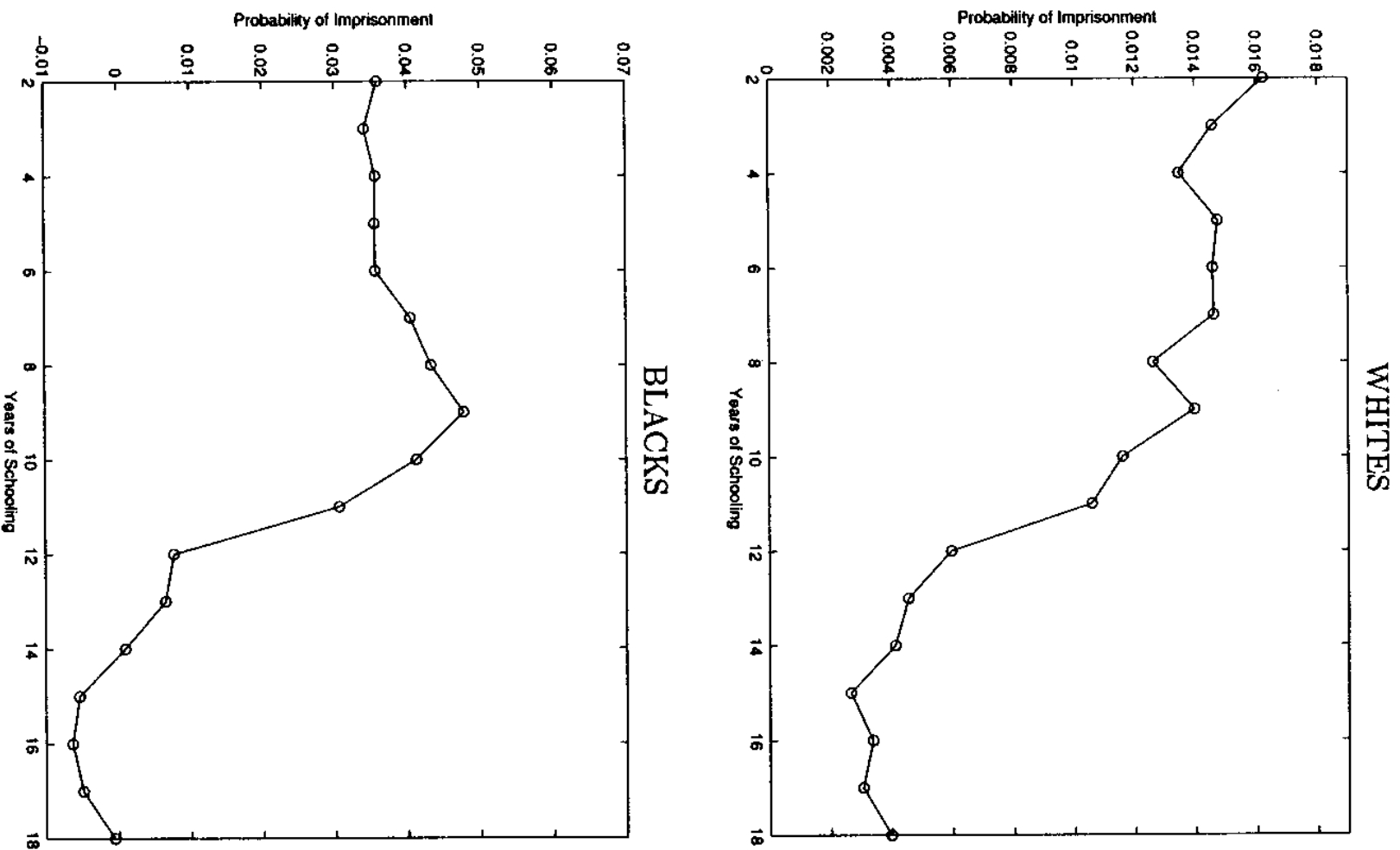
Figure 16

Percentage of Each Gender Who Perform at Level 1 on the IALS Quantitative Literacy Scale



Note: The scale scores were grouped into five levels of increasing difficulty, with Level 1 representing functional illiteracy. Levels 4 and 5 were combined. The sample is restricted to adults who are between 16–65 years of age at the time of the survey (1994 for the US and Germany, 1996 for the UK, and 1994–1995 for Sweden). Standard errors are calculated using the methodology described in IALS (2002).

**Figure 17**  
Regression-Adjusted Probability of  
Incarceration, by Years of Schooling



Source: Lochner and Moretti (2004)

Figure 18

Using the Estimated Technology to Simulate Balanced Interventions

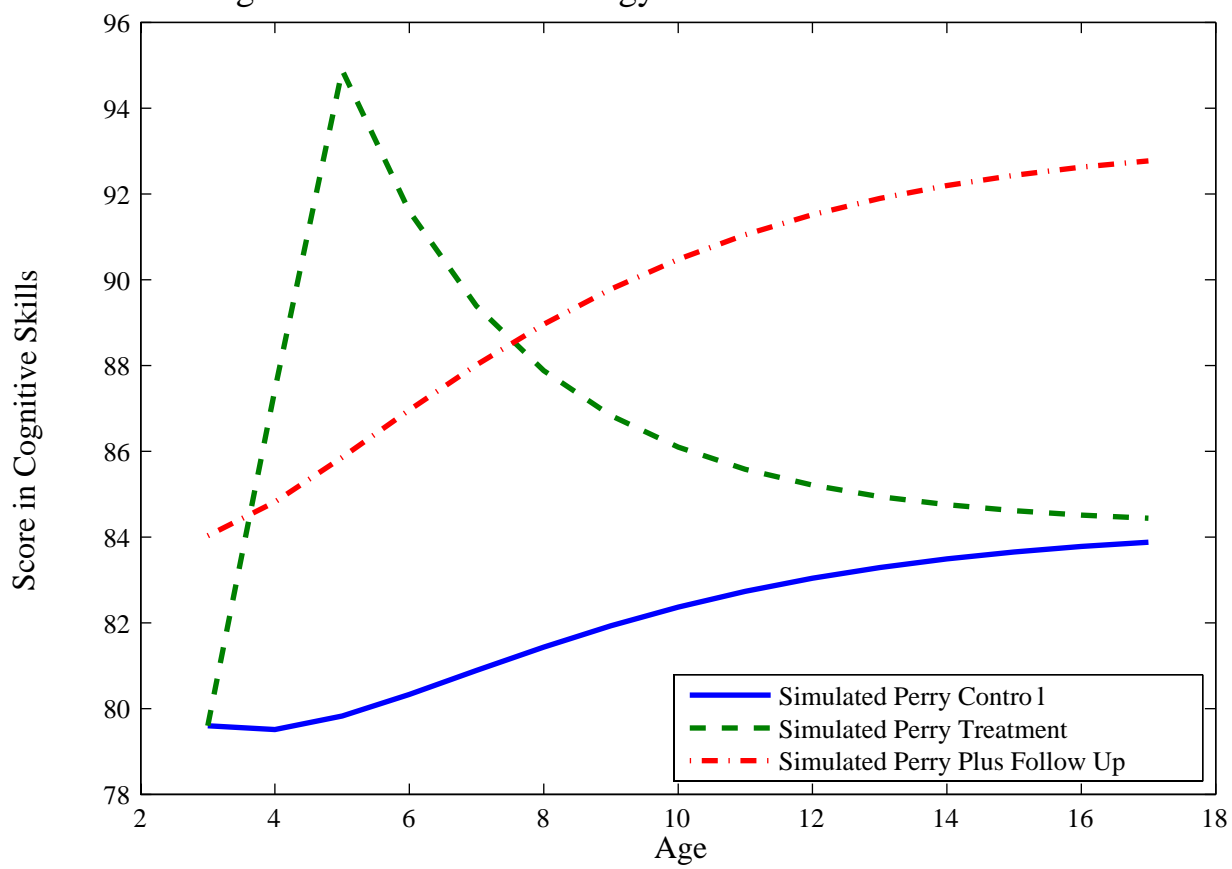
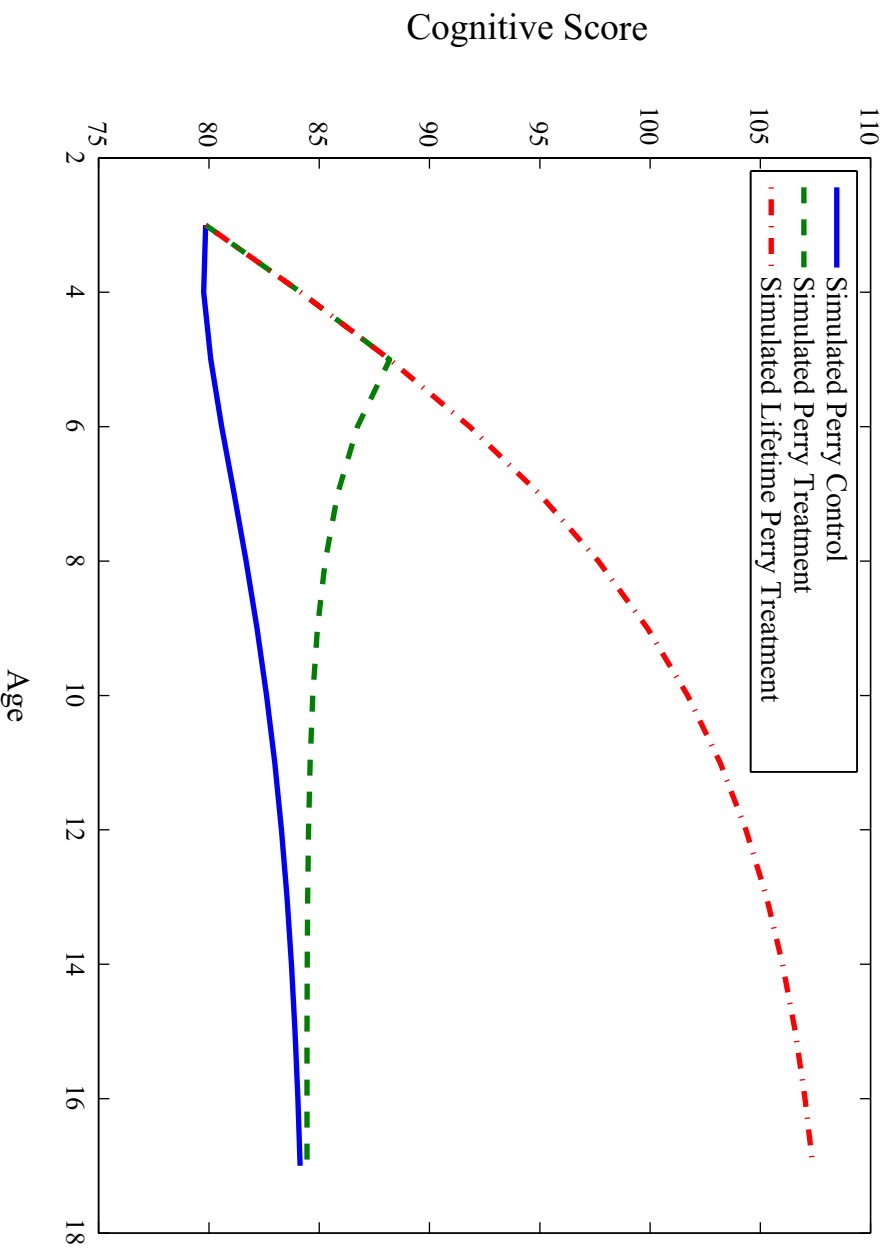


Figure 19  
Using the Estimated Technology to  
Simulate Perry Preschool with Follow Up



**Table 1** Economic Benefits And Costs

	Perry	Chicago CPC
Child Care	986	1,916
Earnings	40,537	32,099
K-12	9,184	5,634
College/Adult	-782	-644
Crime	94,065	15,329
Welfare	355	546
FG Earnings	6,181	4,894
Abuse/Neglect	0	344
Total Benefits	150,525	60,117
Total Costs	16,514	7,738
Net Present Value	134,011	52,380
Benefits-To-Costs Ratio	9.11	7.77

Notes: All values discounted at 3% and are in \$2004. Numbers differ slightly from earlier estimates because FG Earnings for Perry and Chicago were estimated using the ratio of FG Earnings Effect to Earnings Effect (about 15%) that was found in Abecedarian

Source: Barnett, 2004.

**Table 2**

## Effects of Early Intervention Programs

Program/Study	Costs*	Program Description	Test Scores	Schooling	Pre-Delinquency Crime
Abecedarian Project** (Ramey, et. al, 1988)	N/A	full-time year round classes for children from infancy through preschool	high scores at ages 1-4	34% less grade retention by 2nd grade; better reading and math proficiency	
Early Training** (Gray et al., 1982)	N/A	part-time classes for children in summer; weekly home visits during school year	higher scores at ages 5-10	16% less grade retention; 21% higher HS grad.rates	
Harlem Study (Palmer, 1983)	N/A	individual teacher-child sessions twice-weekly for young males	higher scores at ages 3-5	21% less grade retention	
Houston PCDC** (Johnson, 1988)	N/A	home visits for parents for 2 yrs; child nursery care 4 days/wk in year 2 (Mexican Americans)	higher scores at age 3		rated less aggressive and hostile by mothers (ages 8-11)
Milwaukee Project** (Garber, 1988)	N/A	full-time year-round classes for children through 1st grade; job training for mothers	higher scores at ages 2-10	27% less grade retention	

**Table 2 (continued)**

Program/Study	Costs*	Program Description	Test Scores	Schooling	Pre-Delinquency Crime
Mother-Child Home Program (Levenstein, O'Hara, & Madden, 1983)	N/A	home visits with mothers and children twice weekly	higher scores at ages 3-4	6% less grade retention	
Perry Preschool Program** (Schweinhart, Barnes, & Weikart, 1993)	\$13,400	weekly home visits with parents; intensive, high quality preschool services for 1-2 years	higher scores in all studied years (ages 5-27)	21% less grade retention or special services; 21% higher HS grad. rates	2.3 vs. 4.6 lifetime arrests by age 27 7% vs. 35% arrested 5 or more times
Rome Head Start (Monroe & McDonald, 1981)	\$5,400 (2 yrs)	part-time classes for children; parent involvement		12% less grade retention; 17% higher HS grad. rates	
Syracuse University Family Development (Lally et al., 1988)	\$38,100	weekly home visits for family; day care year round	higher scores at ages 3-4		6% vs. 22% had probation files; offenses were less severe
Yale Experiment	\$23,300	family support; home visits and day care as needed for 30 months	better language development at 30 months	better-school attendance & adjustment; fewer special services (age 12 1/2)	rated less aggressive & pre-delinquent by teachers and parents (ages 12 1/2)

Notes: All comparisons are for program participants vs. non-participants. \* Costs valued in 1990 dollars. \*\* Studies used a random assignment experimental design to determine program impacts. Data from Donohue and Siegelman (1998), Schweinhart, Barnes, and Weikart (1993), and Seitz (1990) for the impacts reported here. Source: Heckman, Lochner, Smith, and Taber (1997).

**Table 3**

## Outcomes of Early Intervention Programs

Program (Years of Operation)	Outcome	Followed Up to Age	Age When Treatment Effect Last Statistically Significant	Control Group	Change in Treated Group
<b>Cognitive Measures</b>					
Early Training Project (1962 - 1965)	IQ	16-20	6	82.8	+12.2
Perry Preschool Project (1962 - 1967)	IQ	27	7	87.1	+4.0
Houston PCDC (1970 - 1980)	IQ	8-11	2	90.8	+8.0
Syracuse FDRP (1969 - 1970)	IQ	15	3	90.6	+19.7
Carolina Abecedarian (1972 - 1985)	IQ	21	12	88.4	+5.3
Project CARE (1978 - 1984)	IQ	4.5	3	92.6	+11.6
IHDP (1985 - 1988 )	IQ (HLBW sample)	8	8	92.1	+4.4
<b>Educational Outcomes</b>					
Early Training Project	Special Education	16-20	18	29%	-26%
Perry Preschool Project	Special Education	27	19	28%	-12%
	High School Graduation		27	45%	+21%
Chicago CPC (1967 - present)	Special Education	20	18	25%	-10%
	Grade Retention		15	38%	-15%
	High School Graduation		20	39%	+11%
Carolina Abecedarian	College Enrollment	21	21	14%	+22%
<b>Economic Outcomes</b>					
Perry Preschool Project	Arrest Rate	27	27	69%	-12%
	Employment Rate		27	32%	+18%
	Monthly Earnings		27	\$766	+ \$453
	Welfare Use		27	32%	-17%
Chicago CPC (preschool vs. no preschool)	Juvenile Arrests	20	18	25%	-8%
Syracuse FDRP	Probation Referral	15	15	22%	-16%
Elmira PEIP (1978 - 1982)	Arrests (HR sample)	15	15	0.53	-.029

Notes: HLBW = heavier, low birth weight sample; HR = high risk. Cognitive measures include Stanford-Binet and Wesler Intelligence Scales, California Achievement Tests, and other IQ and achievement tests measuring cognitive ability. All results significant at .05 level or higher. Source: Karoly, 2001. For a discussion of the specific treatments offered under each program see Heckman (2000) and Karoly (2001).

**Table 4.** Anthropomorphic, Developmental, and Cognitive Outcomes Of Romanian and Within-UK Adoptees Over Time

Age of Adoption (Months):	Within-UK Adoptees	Romanian Orphans		
	6	Before 6	At 6-24	At 24-42
Weight at Adoption	-	-2.1	-2.3	-
	-	(1.7)	(1.7)	-
Height at Adoption	-	-1.8	-2.2	-
	-	(1.6)	(2.4)	-
Denver Developmental Scale At Adoption	-	76.5	48.1	-
	-	(48.1)	(25.4)	-
Weight at Age 4	0.45	-0.02	0.04	-
	(0.79)	(0.92)	(0.94)	-
Height at Age 4	0.25	-0.29	-0.36	-
	(0.91)	(0.89)	(1.02)	-
Denver Developmental Scale at Age 4	117.7	115.7	96.7	-
	(24.3)	(23.4)	(21.3)	-
McCarthy GCI at Age 4	109.4	105.9	91.7	-
	(14.8)	(17.9)	(18.0)	-
Weight at Age 6	0.30	0.02	-0.25	-0.85
	(0.90)	(0.97)	(0.96)	(0.98)
Percentage With Denver Developmental Scale at Age 6 Below 70	2	0	5	18
	(1)	(0)	(2)	(7)
McCarthy GCI at Age 6	117	114	99	90
	(17.8)	(18.3)	(19.2)	(23.8)

Standard deviations are reported below in parentheses. All anthropometric measurements are standardized using the UK age-specific distributions. The Denver Developmental Scale is based on specific behaviors (*e.g.*, standing while holding on to something, lifting the head, making meaningful "da-da" sounds). Due to ceiling effects, the Denver scale is not meaningful at age 6, so O'Connor *et al.* (2000) use the percentage with impairment (defined as a score below 70) as the test criterion. The GCI is the total score on the McCarthy Scales of Children's Abilities. It summarizes verbal, quantitative, perceptual, and memory performance. See Rutter *et al.* (1998) and O'Connor *et al.* (2000) for more details on the analysis.

**Table 5**

Estimated Benefits of Mentoring Programs (Treatment Group Reductions Compared to Control Group)

Program	Outcome Measure	Change	Program Costs per Participant
Big Brother / Big Sister	Initiating drug use	-45.8%	\$500 - \$1500*
	Initiation alcohol use	-27.4%	
	# of times hit someone	-31.7%	
	# of times stole something	-19.2%	
	Grade Point Average	3.0%	
	Skipped Class	-36.7%	
	Skipped Day of School	-52.2%	
	Trust in Parent	2.7%	
	Lying to Parent	-36.6%	
	Peer Emotional Support	2.3%	
Sponsor - A - Scholar			\$1485
	10th Grade GPA (100 point scale)	2.9	
	11th Grade GPA (100 point scale)	2.5	
	% Attending College (1 year after HS)	32.8%	
	% Attending College (2 years after HS)	28.1%	
Quantum Opportunity Program	Graduated HS or GED	+26%	
	Enrolled in 4-year college	+15%	
	Enrolled in 2-year college	+24%	
	Currently employed full time	+13%	
	Self receiving welfare	-22%	
	% ever arrested	-4%	

Sources: Benefits from Heckman (1999) and Taggart (1995), costs from Johnson (1996) and Herrera et al. (2000).

Notes: \*Costs, in 1996 dollars, for school-based programs are as low as \$500 and more expensive community based mentoring programs cost as high as \$1500; HS = high school

**Table 6**

Effects of Selected Adolescent Social Programs on Schooling, Earnings, and Crime

Program/Study	Costs*	Program Description	Schooling	Earnings*	Crime*
Job Corps (Long et al., 1981)	\$11,000	7 mo. of educ. and vocational training for 16-21 yr. olds (mostly male)	no effect	disc. pres. value of increased earnings of \$10,000	Estimated Reduction in crime valued at approx.
STEP (Walker and Viella-Velez, 1992)	N/A	2 summers of employment, academic remediation & life skills for 14 & 15 year olds	short-run gains in test scores; no effect on school completion rates		
Quantum Opportunities Program** (Taggart, 1995)	\$10,600	counseling; educ., comm., & devp. services; financial incentives for part. (4 yrs. beginning in 9th grade)	34% higher HS grad./GED rates (2 yrs. post-program)		4% vs. 16% convicted; .28 vs. .56 avg. number of arrests (2 yrs. post- program)

**Table 7.** The Ratio of Optimal Early and Late Investments  $\frac{I_1}{I_2}$  Under Different Assumptions About the Skill Formation Technology

	Low Self-Productivity: $\gamma < \frac{(1+r)}{(2+r)}$	High Self-Productivity: $\gamma > \frac{(1+r)}{(2+r)}$
High Degree of Complementarity: $\phi < 0$	$\frac{I_1}{I_2} \rightarrow 1$ as $\phi \rightarrow -\infty$	$\frac{I_1}{I_2} \rightarrow 1$ as $\phi \rightarrow -\infty$
Low Degree of Complementarity: $0 \leq \phi \leq 1$	$\frac{I_1}{I_2} \rightarrow 0$ as $\phi \rightarrow 1$	$\frac{I_1}{I_2} \rightarrow \infty$ as $\phi \rightarrow 1$

*Note:* This table summarizes the behavior of the ratio of optimal early to late investments according to four cases:  $I_1$  and  $I_2$  have high complementarity, but self-productivity is low;  $I_1$  and  $I_2$  have both high complementarity and self-productivity;  $I_1$  and  $I_2$  have low complementarity and self-productivity; and  $I_1$  and  $I_2$  have low complementarity, but high self-productivity. When  $I_1$  and  $I_2$  exhibit high complementary, complementarity dominates and is a force towards equal distribution of investments between early and late periods. Consequently, self-productivity plays a limited role in determining the ratio  $\frac{I_1}{I_2}$  (row 1). On the other hand, when  $I_1$  and  $I_2$  exhibit a low degree of complementarity, self-productivity tends to concentrate investments in the *late* period if self-productivity is low, but in the *early* period if it is high (row 2).

**Table 8A: Summary Dynamic Measurements - White Children NLSY/1979**

	Obs	Age 6		Age 7		Age 8		Age 9				
		Mean	Std Error	Mean	Std Error	Mean	Std Error	Mean	Std Error			
Piat Math <sup>1</sup>	460	-1.2174	0.4063	358	-0.7701	0.5407	493	-0.0953	0.5995	374	0.2776	0.5952
Piat Reading Recognition <sup>1</sup>	456	-1.2363	0.2874	360	-0.8312	0.4689	490	-0.2128	0.6235	373	0.1011	0.6626
Piat Composition <sup>1</sup>	430	-1.2748	0.2757	331	-0.8583	0.5008	469	-0.1932	0.6880	369	0.1384	0.6459
Antisocial Score <sup>1</sup>	453	0.1017	0.9962	363	0.0853	0.9506	490	-0.0262	1.0420	379	-0.0948	1.0872
Anxious Score <sup>1</sup>	471	0.2302	0.9931	371	0.1065	1.0019	496	-0.0239	1.0305	385	-0.0340	1.0126
Headstrong Score <sup>1</sup>	471	0.0291	0.9853	373	0.0451	0.9892	496	-0.1822	1.0044	385	-0.2089	1.0126
Hyperactive Score <sup>1</sup>	472	-0.0803	0.9771	373	-0.0392	0.9515	497	-0.1505	0.9899	386	-0.0404	1.0245
Conflict Score <sup>1</sup>	471	0.0238	1.0154	372	0.0452	0.9555	498	0.0494	0.9631	385	-0.0335	1.0153
Log Current Family Income <sup>2</sup>	674	10.3480	1.3756	684	10.3804	1.2097	702	10.2735	1.6936	687	10.2073	1.7313
Number of Books <sup>3</sup>	321	3.9221	0.3310	373	3.9249	0.3513	502	3.9522	0.2315	387	3.8941	0.3693
Musical Instrument <sup>4</sup>	320	0.4750	0.5002	373	0.4879	0.5005	503	0.4612	0.4990	386	0.5596	0.4971
Newspaper <sup>4</sup>	321	0.5234	0.5002	373	0.5496	0.4982	503	0.5129	0.5003	386	0.4922	0.5006
Child has special lessons <sup>4</sup>	321	0.5265	0.5001	371	0.5957	0.4914	502	0.6853	0.4649	386	0.7409	0.4387
Child goes to museums <sup>5</sup>	320	2.2438	0.9452	373	2.2949	0.8853	503	2.2843	0.8258	386	2.3290	0.8577
Child goes to theater <sup>5</sup>	322	1.7578	0.7915	372	1.8629	0.8531	502	1.7729	0.7370	386	1.8601	0.7804
Child ever sees father <sup>4</sup>	284	0.9754	0.1553	308	0.9870	0.1134	445	0.9820	0.1330	330	0.9788	0.1443
Child spends time with father indoors <sup>6</sup>	318	4.9340	1.5295	370	4.8892	1.6564	490	4.9959	1.4329	381	5.1811	1.3577
Child spends time with father outdoors <sup>6</sup>	316	4.3734	1.2188	364	4.4396	1.2832	491	4.2892	1.1788	371	4.4528	1.2125
Child eats with father and mother <sup>6</sup>	312	4.5801	1.3005	368	4.7663	1.2241	489	4.5869	1.3140	377	4.6923	1.2593
Child sees relatives and family friends <sup>6</sup>	321	3.9190	1.1591	371	3.8518	1.1704	502	3.7450	1.1949	387	3.6718	1.2272

<sup>1</sup>The variables are standardized with mean zero and variance one across the entire CNLSY/79 sample.

<sup>2</sup>Family Income is inflation adjusted. Base year is 2000.

<sup>3</sup>The variable takes the value 1 if the child has no books, 2 if the child has 1 or 2 books, 3 if the child has 3 to 9 books and 4 if the child has 10 or more books.

<sup>4</sup>For example, for musical instrument, the variable takes value 1 if the child has a musical instrument at home and 0 otherwise. Other variables are defined accordingly.

<sup>5</sup>For example, for "museums", the variable takes the value 1 if the child never went to the museum in the last calendar year, 1 if the child went to the museum once or twice in the last calendar year, 3 if the child went to the museum several times in the past calendar year, 4 if the child went to the museum about once a month in the last calendar year, and 5 if the child went to a museum once a week in the last calendar year.

<sup>6</sup>For example, for "Child spends time with father indoors", the variable takes the value 1 if the child never spends time with the father indoors, 2 if the child spends time with the father indoors a few times in a year, 3 if the child spend time with the father indoors about once a month, 4 if the child spends time with the father indoors about once a week, 5 if the child spends time with the father indoors at least four times a week, and 6 if the child spends time with the father once a day or more often.

**Table 9**  
**The Technology Equations**  
**Measurement Variables are Standardize with Mean Zero and Variance One**  
**Unanchored Model**

	Symbol	Next Period Noncognitive Skills		Next Period Cognitive Skills	
		Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	$\gamma_{k,1}$	0.8998	0.0212	0.0334	0.0122
Current Period Cognitive Skills	$\gamma_{k,2}$	0.0201	0.0123	0.8974	0.0312
Current Period Investment	$\gamma_{k,3}$	0.0654	0.0211	0.0371	0.0122
Mother's Cognitive Skill	$\gamma_{k,4}$	0.0002	0.0082	0.0321	0.0066
Mother's Noncognitive Skill	$\gamma_{k,5}$	0.0144	0.0072	0.0000	0.0013

Source: Cunha and Heckman (2006)

---

---

**Table 10****Per Period Correlation Matrices**

Period 1 - Children ages 6 and 7

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2087	0.2899
Cognitive	0.2087	1.0000	0.2615
Investment (Home)	0.2899	0.2615	1.0000

Period 2 - Children ages 8 and 9

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2399	0.3526
Cognitive	0.2399	1.0000	0.2908
Investment (Home)	0.3526	0.2908	1.0000

Period 3 - Children ages 10 and 11

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2657	0.4117
Cognitive	0.2657	1.0000	0.3258
Investment (Home)	0.4117	0.3258	1.0000

Period 4 - Children ages 12 and 13

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2851	0.4551
Cognitive	0.2851	1.0000	0.3539
Investment (Home)	0.4551	0.3539	1.0000

---

Source: Cunha and Heckman (2006)

**Table 11**  
**The Technology Equations<sup>1</sup>**  
**Measurement Variables are Standardize with Mean Zero and Variance One**  
**Results Using Linear Probability Model for High-School Graduation**

	Symbol	Next Period Noncognitive Skills		Next Period Cognitive Skills	
		Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	$\gamma_{k,1}$	0.8998	0.0212	0.0388	0.0176
Current Period Cognitive Skills	$\gamma_{k,2}$	0.0146	0.0153	0.8974	0.0312
Current Period Investment	$\gamma_{k,3}$	0.0443	0.0294	0.0523	0.0108
Mother's Cognitive Skill	$\gamma_{k,4}$	0.0000	0.0138	0.0115	0.0077
Mother's Noncognitive Skill	$\gamma_{k,5}$	0.0413	0.0088	0.0000	0.0025

**Table 12**  
**Stage Specific Technology Parameters**  
**The Technology Equations**  
**Measurement Variables are Standardize with Mean Zero and Variance One**  
**Results Using Linear Probability Model for High-School Graduation**

		Non-Cognitive Skills		
Dependent Variable	Symbol	Transition 1*	Transition 2*	Transition 3*
Current Period Non-Cognitive Skills	$\gamma_{t,N,1}$	0.9345**	0.9147**	0.8582**
Current Period Cognitive Skills	$\gamma_{t,N,2}$	0.0301	0.0310	0.0379
Parental Investment	$\gamma_{t,N,3}$	0.0204**	0.0593**	0.1038**
Mother's Cognitive Skills	$\gamma_{t,N,4}$	0.0002	0.0004	0.0000
Mother's Non-Cognitive Skills	$\gamma_{t,N,5}$	0.0149**	0.0246	0.0001
		Cognitive Skills		
	Symbol	Transition 1*	Transition 2*	Transition 3*
Current Period Non-Cognitive Skills	$\gamma_{t,C,1}$	0.0662**	0.0240**	0.0000
Current Period Cognitive Skills	$\gamma_{t,C,2}$	0.8171**	0.9318	0.9630**
Parental Investment	$\gamma_{t,C,3}$	0.0720**	0.0178**	0.0160**
Mother's Cognitive Skills	$\gamma_{t,C,4}$	0.0447**	0.0231**	0.0217**
Mother's Non-Cognitive Skills	$\gamma_{t,C,5}$	0.0000	0.0000	0.0000

Source: Cunha and Heckman (2006)

**Table 13****The Weights in the Construction of the Investment Factor<sup>1</sup>**

Ages 6 and 7

Ages 8 and 9

	Estimated Weights	Ad Hoc Weights	Share of Total Residual	Share of Total Residual	Estimated Weights	Ad Hoc Weights	Share of Total Residual	Share of Total Residual
			Variance due to Factors	Variance due to Uniqueness			Variance due to Factors	Variance due to Uniqueness
Number of Books	0.3079	0.1667	0.1242	0.8758	0.2971	0.1667	0.0943	0.9057
Musical Instrument	0.1997	0.1667	0.1417	0.8583	0.1667	0.1667	0.1312	0.8688
Newspaper	0.1932	0.1667	0.1517	0.8483	0.2148	0.1667	0.0766	0.9234
Child has special lessons	0.1431	0.1667	0.2808	0.7192	0.1560	0.1667	0.1801	0.8199
Child goes to museums	0.0740	0.1667	0.3063	0.6937	0.0768	0.1667	0.2158	0.7842
Child goes to theater	0.0821	0.1667	0.3068	0.6932	0.0821	0.1667	0.2473	0.7527

Ages 10 and 11

Ages 12 and 13

	Estimated Weights	Ad Hoc Weights	Share of Total Residual	Share of Total Residual	Estimated Weights	Ad Hoc Weights	Share of Total Residual	Share of Total Residual
			Variance due to Factors	Variance due to Uniqueness			Variance due to Factors	Variance due to Uniqueness
Number of Books	0.1396	0.1667	0.1056	0.8944	0.1380	0.1667	0.0568	0.9432
Musical Instrument	0.1707	0.1667	0.1196	0.8804	0.2001	0.1667	0.0500	0.9500
Newspaper	0.2082	0.1667	0.0754	0.9246	0.2253	0.1667	0.0370	0.9630
Child has special lessons	0.1971	0.1667	0.1074	0.8926	0.2396	0.1667	0.0457	0.9543
Child goes to museums	0.0866	0.1667	0.1693	0.8307	0.0910	0.1667	0.1011	0.8989
Child goes to theater	0.0821	0.1667	0.2074	0.7926	0.0821	0.1667	0.1412	0.8588

**Table 14**  
**The Nonlinear Technology Equations<sup>1</sup>**  
**Unanchored Model**

	Symbol	Next Period Noncognitive Skills		Next Period Cognitive Skills	
		Mean	Standard Error	Mean	Standard Error
Constant	$B_k$	0.6932	0.0374	1.0541	0.0834
Current Period Noncognitive Skills	$\gamma_{k,1}$	0.7912	0.0297	0.0213	0.0103
Current Period Cognitive Skills	$\gamma_{k,2}$	0.0372	0.0178	0.8673	0.0423
Current Period Investments	$\gamma_{k,3}$	0.0828	0.0269	0.0599	0.0217
Mother's Cognitive Skills	$\gamma_{k,4}$	0.0250	0.0105	0.0314	0.0139
Current Period Noncognitive Skills	$\gamma_{k,5}$	0.0639	0.0207	0.0201	0.0102
Parameter of the Elasticity of Substitution	$\phi_k$	-0.1710	0.0322	-0.8961	0.0763

**Table 15**  
**The Nonlinear Technology Equations<sup>1</sup>**  
**Anchoring on the Probability of Graduating from High School using a Probit Model**

	Symbol	Next Period Noncognitive Skills		Next Period Cognitive Skills	
		Mean	Standard Error	Mean	Standard Error
Constant	$B_k$	1.4226	0.0484	0.9842	0.0932
Current Period Noncognitive Skills	$\gamma_{k,1}$	0.7403	0.0359	0.0455	0.0133
Current Period Cognitive Skills	$\gamma_{k,2}$	0.0516	0.0234	0.7206	0.0581
Current Period Investment	$\gamma_{k,3}$	0.1262	0.0302	0.1168	0.0384
Mother's Cognitive Skills	$\gamma_{k,4}$	0.0151	0.0178	0.0724	0.0265
Mother's Noncognitive Skills	$\gamma_{k,5}$	0.0668	0.0269	0.0446	0.0153
Substitution	$\phi_k$	-0.1234	0.0419	-0.2543	0.0839

**Table 16**  
**Characteristics of the Labor Force Aged 25 and Over**  
**and Components of Change 1980, 2000, 2020**  
**(Millions of Workers)**

	Labor Force in 1980	<b>Growth 1980-2000</b>	Labor Force in 2000	<b>Growth 2000-2020</b>	Labor Force in 2020
<b>Age</b>					
25-54	65.0	35.1	100.1	3.0	103.1
55-64	11.8	2.2	14.0	12.5	26.5
65+	3.0	1.4	4.4	4.0	8.4
<i>Total</i>	79.8	38.7	118.5	19.4	137.9
<b>Race/Ethnicity/Nativity</b>					
White Non-Hispanic – Native	63.0	21.5	84.5	2.6	87.1
Black Non-Hispanic – Native	7.6	4.6	12.2	2.8	15.0
Hispanic – Native	2.5	2.3	4.8	6.8	11.6
Other Non-Hispanic – Native	0.8	1.0	1.8	1.2	3.0
Hispanic – Foreign Born	1.8	4.5	6.3	2.8	9.1
Non-Hispanic – Foreign Born	4.1	4.8	8.9	3.3	12.2
<i>Total</i>	79.8	38.7	118.5	19.4	137.9
<b>SUMMARY</b>					
Native White Workers 25-54	50.8	19.3	70.1	-7.7	62.4
Native White Workers 55 & Over	12.2	2.2	14.4	10.3	24.7
Workers of Color 25-54	9.4	7.3	16.7	7.7	24.4
Workers of Color 55 & Over	1.6	0.5	2.1	3.0	5.1
Foreign Born Workers	5.9	9.4	15.3	6.0	21.3
<i>Total</i>	79.8	38.7	118.5	19.4	137.9

Source: Ellwood (2001).

**Table 17**  
**Educational Characteristics of the Labor Force**  
**Aged 25 and Over**  
**1980, 2000, 2020**

	Labor Force in 1980	<b>Growth 1980-2000</b>	Labor Force in 2000	<b>Growth 2000-2020</b>	Labor Force in 2020
<b>Education</b>					
Less Than High School	17.3	-5.3	12.0	0.9	12.9
High School Only	31.5	6.3	37.8	3.8	41.6
Some Schooling Beyond High School	13.8	19.1	32.9	6.2	39.1
College Degree or More	17.3	18.5	35.8	7.7	43.5
<b>Total</b>	<b>79.8</b>	<b>38.7</b>	<b>118.5</b>	<b>18.6</b>	<b>137.1</b>
% With College Degree	21.6%		30.2%		31.7%

\*Assumes that subsequent cohorts have same education at age 25 as the cohort age 25 in 2000.

Source: Ellwood (2001).

**Table 18.** Aggregate Burden Of Crime

Crime-induced Production (\$ billion)	464
Opportunity Costs (\$ billion)	152
Risks to Life And Health (\$ billion)	672
Transfers (\$ billion)	706
Gross Burden (\$ billion)	1,995
Net of Transfers (\$ billion)	1,289
Per Capita (\$)	4,818

Source: Anderson (1999). All figures inflated to \$2004 using the CPI.

**Table 19 .** Estimated Social Benefits Of Increasing High School

Completion Rates By 1 Percent		
	Estimated Change In Crime	Social Benefits
<b>Violent Crimes:</b>		
Murder	-373	\$1,457,179,565
Rape	1,559	-\$179,450,969
Robbery	918	-\$11,116,176
Assault	-37,135	\$475,045,373
<b>Property Crimes:</b>		
Burglary	-9,467	\$12,052,009
Larceny/Theft	-35,105	\$8,958,962
Motor Vehicle Theft	-14,238	\$22,869,192
Arson	-469	\$23,637,635
<b>Total:</b>	<b>-94,310</b>	<b>\$1,809,175,590</b>

Notes: Victim costs and property losses taken from Table 2 of Miller *et al.* (1996). Incarceration costs per crime equal the incarceration cost per inmate, \$17,027 (U.S. Department of Justice, 1999), multiplied by the incarceration rate (U.S. Department of Justice, 1994). Total costs are calculated as the sum of victim costs and incarceration costs less 80% of the property loss (already included in victim costs) for all crimes except arson. Total costs for arson are the sum of victim costs and incarceration costs since there is no transfer of property between victim and criminal. Estimated changes in crimes adjusts the arrest effect by the number of crimes per arrest. The social benefits is the estimated change in crimes times the total cost per crime. All dollar figures are adjusted to \$2004 using the CPI. Source: Lochner and Moretti (2004).

**Table 20**  
**Comparison of Different Investment Strategies**  
**Disadvantaged Children: First Decile in the Distribution of Cognitive and Non-Cognitive Skills at Age 6**  
**Mothers are in First Decile in the Distribution of Cognitive and Non-Cognitive Skills at Ages 14-21**

	Baseline	Changing Initial Conditions - Moving Children to the 4 <sup>th</sup> Decile of Distribution of Skills only through Early Investment	Adolescent Intervention: Moving Investments at Last Transition from 1st to 9th Decile
High School Graduation	0.4109	0.6579	0.6391
Enrollment in College	0.0448	0.1264	0.1165
Conviction	0.2276	0.1710	0.1773
Probation	0.2152	0.1487	0.1562
Welfare	0.1767	0.0905	0.0968

Table 21  
Comparison of Different Investment Strategies

Disadvantaged Children: First Decile in the Distribution of Cognitive and Non-Cognitive Skills at Age 6  
Mothers are in First Decile in the Distribution of Cognitive and Non-Cognitive Skills at Ages 14-21

	Baseline	Changing Initial Conditions - Moving Children to the 4 <sup>th</sup> Decile of Distribution of Skills	Changing Initial Conditions and Performing Adolescent Intervention	Changing Initial Conditions and Performing a Balanced Intervention
High School Graduation	0.4109	0.6579	0.8477	0.9135
Enrollment in College	0.0448	0.1264	0.2724	0.3755
Conviction	0.2276	0.1710	0.1272	0.1083
Probation	0.2152	0.1487	0.1009	0.0815
Welfare	0.1767	0.0905	0.0415	0.0259