NBER WORKING PAPER SERIES

FOSTERING AND MEASURING SKILLS: INTERVENTIONS THAT IMPROVE CHARACTER AND COGNITION

James J. Heckman
Tim Kautz

Working Paper 19656
http://www.nber.org/papers/w19656

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2013

This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, the Buffett Early Childhood Fund, NICHD 5R37HD065072, 5R01HD054702, the Human Capital and Economic Opportunity Global Working Group--an initiative of the Becker Friedman Institute for Research in Economics--funded by the Institute for New Economic Thinking (INET), and an anonymous funder. We acknowledge the support of an European Research Council grant hosted by the University College Dublin, DEVHEALTH 269874. The views expressed in this paper are those of the authors and not necessarily those of the funders or commentators mentioned here. We thank Linor Kiknadze and Edward Sung for valuable research assistance. We received helpful comments from Richard Boyle, Ron Diris, Maryclare Grin, Robert Lerman, Seong Hyeok Moon, Dan Moran, Maria Rosales, and Indra Wechsberg. A version of this paper appears in Heckman, Humphries, and Kautz (2014b): see Heckman and Kautz (2014b). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by James J. Heckman and Tim Kautz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Fostering and Measuring Skills: Interventions That Improve Character and Cognition
James J. Heckman and Tim Kautz
NBER Working Paper No. 19656
November 2013
JEL No. D01,I20,J24

ABSTRACT

This paper reviews the recent literature on measuring and boosting cognitive and noncognitive skills. The literature establishes that achievement tests do not adequately capture character skills--personality traits, goals, motivations, and preferences--that are valued in the labor market, in school, and in many other domains. Their predictive power rivals that of cognitive skills. Reliable measures of character have been developed. All measures of character and cognition are measures of performance on some task. In order to reliably estimate skills from tasks, it is necessary to standardize for incentives, effort, and other skills when measuring any particular skill.

Character is a skill, not a trait. At any age, character skills are stable across different tasks, but skills can change over the life cycle. Character is shaped by families, schools, and social environments. Skill development is a dynamic process, in which the early years lay the foundation for successful investment in later years.

High-quality early childhood and elementary school programs improve character skills in a lasting and cost-effective way. Many of them beneficially affect later-life outcomes without improving cognition. There are fewer long-term evaluations of adolescent interventions, but workplace-based programs that teach character skills are promising. The common feature of successful interventions across all stages of the life cycle through adulthood is that they promote attachment and provide a secure base for exploration and learning for the child. Successful interventions emulate the mentoring environments offered by successful families.

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and University College Dublin and IZA
and also NBER
jjh@uchicago.edu

Tim Kautz
University of Chicago
Department of Economics
1126 E. 59th Street
Chicago IL 60637
tkautz@uchicago.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w19656

Contents

1 Introduction **3**

2 Measures of Cognitive and Character Skills **9**

2.1 Cognitive Skills 9

2.2 Measuring Character 10

2.3 A Task-Based Framework for Identifying and Measuring Skills 13

2.4 Reference Bias 18

2.5 Measuring Skills Using Behaviors 20

2.6 Are Character Skills Stable? 22

3 Predictive Power of Character **23**

4 The Skills Needed for Success in the Labor Market **29**

5 A Framework for Understanding Interventions **31**

6 Summary of Empirical Evidence on the Efficacy of Interventions **33**

7 Early-Life Interventions that Begin Before Formal Schooling **36**

7.1 Infant Programs and Model Preschools 36

7.1.1 Nurse-Family Partnership 36

7.1.2 Jamaican Study 41

7.1.3 Perry Preschool Program 43

7.1.4 Abecedarian Program 47

7.2 Large-Scale Infant and Young Child Programs 49

7.2.1 Head Start 49

7.2.2 Chicago CPC 50

8 Education and Interventions in Kindergarten and Elementary School **52**

8.1 Targeted Character Interventions 52
8.1.1 The Seattle Social Development Project (SSDP) 53
8.1.2 Cambridge-Somerville Program 56
8.1.3 Project STAR 56

9 Education and Interventions Targeted Toward Adolescents and Young Adults 57

9.1 Adolescent Mentorship Programs 58
9.1.1 Quantum Opportunity Program 58
9.1.2 Big Brothers Big Sisters 60
9.1.3 EPIS Program 62
9.2 Residential-Based Programs 62
9.2.1 Job Corps 62
9.2.2 National Guard ChalleNGe 64
9.3 Workplace-Based Adolescent Intervention Programs 65
9.3.1 Career Academies 67
9.3.2 Year-Up Program 69
9.3.3 Dominican Youth Employment Program 71
9.3.4 Self-Sufficiency Project 72

10 Apprenticeship Programs 74

11 Other Curricula that Have Been Applied to Multiple Age Groups 78

11.1 Tools of the Mind 78
11.2 Studies that Teach the Incremental Theory of Intelligence 78

12 The Effects of Education and Parental Investment on Character and Cognition 81

13 Summary 88

1 Introduction

Modern societies rely on written tests. Achievement tests have come to play an especially prominent role. They are used to sift and sort people, to evaluate schools, and to assess the performance of entire nations.[1] The No Child Left Behind Act requires that public schools administer achievement tests and that the test results influence local school policy. In the US, high school dropouts can take a 7-and-a-half hour achievement test—the General Educational Development (GED) exam—to certify that they are equivalent to high school graduates.[2]

Despite their widespread use, achievement tests are not well understood. Achievement tests were developed in the mid-twentieth century as a way to measure "*general knowledge*" that would be useful inside and outside of the classroom.[3] Their developers claimed to have designed pencil-and-paper tests that would predict success in the labor market, in education, and in many other aspects of life. However, achievement tests are typically validated in a circular fashion using IQ tests and grades, and not in terms of their ability to predict important life outcomes. Their validity in predicting success in outcomes that matter is not well established.

Achievement test scores predict only a small fraction of the variance in later-life success. For example, adolescent achievement test scores only explain about 15% of the variance in later-life earnings.[4] It is unlikely that measurement error accounts for most of the remaining variance.[5] Something fundamental is missing.

Achievement tests do not adequately capture *character skills* such as conscientiousness,

[1]The Programme for International Student Assessment (PISA) evaluates student performance in math, science, and reading across countries, and its results attract a lot of media attention and influence policy. Scores from the year 2000 PISA test led Germany to reevaluate its educational system and introduce a variety of reforms (Grek, 2009).

[2]See Heckman et al. (2014b) for a detailed discussion of the GED program and an evaluation of its benefits.

[3]For histories of achievement tests see Heckman and Kautz (2014a); Quinn (2014).

[4]See Heckman and Kautz (2012).

[5]See Bound, Brown, and Mathiowetz (2001). At most 25%–30% of the variance in hourly wages is due to measurement error.

perseverance, sociability, and curiosity, which are valued in the labor market, in school, and in many other domains. Until recently these skills have largely been ignored. However, in recent research economists and psychologists have constructed measures of these skills and provide evidence that they are stable across situations and predict meaningful life outcomes.[6]

Skills are not set in stone at birth. They can be improved. Cognitive and character skills change with age and with instruction. Interventions to improve skills are effective to different degrees for different skills at different ages. Importantly, character skills are more malleable at later ages.

This paper reviews the recent evidence from economics and personality psychology on the predictive power of cognition and character and the best available evidence on how to foster them. A growing body of empirical research shows that character skills rival IQ in predicting educational attainment, labor market success, health, and criminality.[7]

Character skills are universally valued across all cultures and societies. Recognition of the importance of skills other than raw intelligence is deeply embedded in folk wisdom. Children everywhere are taught character-building stories like *The Tortoise and the Hare* and *The Little Engine That Could*. Even the enthusiastic creators of the early IQ tests, such as Alfred Binet, Charles Spearman, and Edward Webb, recognized the importance of character skills beyond cognition in predicting academic success.[8]

Because both cognition and character can be shaped, and change over the life cycle, we refer to them as "skills" throughout this paper. An older terminology refers to them as "traits," conveying a sense of immutability or permanence, possibly due to their heritable nature. Our distinction between skills and traits is not just a matter of semantics. It suggests

[6]See the studies by Borghans, Duckworth, Heckman, and ter Weel (2008) and Almlund, Duckworth, Heckman, and Kautz (2011). The modern literature traces back to Bowles and Gintis (1976), and Bowles, Gintis, and Osborne (2001). An important study in sociology is the work of Peter Mueser reported in Jencks (1979). Work in psychology going back to Terman, Baldwin, Bronson, DeVoss, Fuller, Lee Kelley, Lima, Marshall, Moore, Raubenheimer, Ruch, Willoughby, Benson Wyman, and Hazeltine Yates (1925) shows that personality skills predict life outcomes (see also Murray, 1938; Terman, Oden, Bayley, Marshall, McNemar, and Sullivan, 1947; and the analysis in Gensowski, 2012).

[7]See Heckman and Kautz (2012), Almlund et al. (2011), Borghans et al. (2008), and Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007) for reviews.

[8]Binet and Simon (1916); Webb (1915).

new and productive avenues for public policy.

Skills enable people. They are capacities to function. Greater levels of skill foster social inclusion and promote economic and social mobility. They generate economic productivity and create social well-being. Skills give agency to people to shape their lives in the present and to create future skills.

An effective strategy for promoting human development should be based on three factually based insights:

1. The powerful role of families in shaping skills;

2. The multiplicity of skills required for successful functioning in society. A core set of skills promotes success in many aspects of life. Different tasks require different skills in different levels and proportions. People tend to pursue the tasks where their skills give them comparative advantage;

3. The technology of skill formation: that skills together with investment beget further skills.

Effective policies to promote skills straddle the missions of cabinet agencies and draw on the wisdom of many academic disciplines. They require broad thinking. Both cognition and character are important ingredients of successful lives. They are malleable to different degrees at different stages of the life cycle. They cross-fertilize each other. Focusing on one dimension of human skills to the exclusion of other dimensions misses fundamental aspects of human performance and development. Narrowly focused policies fail to capture synergisms in the expression and development of skills.

Many policymakers share a common desire to develop human potential. However, current policy discussions focus on promoting skills by improving schools. In this very narrow view, the success of schools is measured by scores on exams used to monitor performance under No Child Left Behind. This focus is a consequence of a very limited conceptualization of human capabilities that assumes that achievement tests capture the important life skills.

This emphasis misses important dimensions of human flourishing. It does not recognize that skills are multiple in nature. Nor does it recognize the importance of families and communities in creating skills. While schools are important, they are not the sole producers of the skills that matter.

Both cognitive and character skills are crucial to success in economic and social life. Character skills include perseverance ("grit"), self-control, trust, attentiveness, self-esteem and self-efficacy, resilience to adversity, openness to experience, empathy, humility, tolerance of diverse opinions, and the ability to engage productively in society.

Our emphasis on character skills does not arise from any agenda to impose Western middle-class values on society. A strong base of cognitive and character skills is universally valued across different cultures, religions, and societies. There are reliable ways to measure them, and there are proven ways to enhance them and to evaluate efforts to foster them.

Recent research establishes the existence of critical and sensitive periods in the formation of skills over the life cycle. Sensitive periods are those when investment is especially productive; critical periods are those when investment is essential. Critical and sensitive periods differ across skills. Investments should target those periods.

In designing effective human development strategies, it is essential to discard obsolete views about the origin and malleability of "traits." What used to be regarded as traits fixed at conception are now understood to be skills that can be augmented through guidance and instruction. Raw intelligence is not fixed solely by parental genes, although heritability plays an important role. It is boosted by quality parenting and by caring environments. It becomes solidified around the time of puberty. Character skills can also be shaped. Compared to raw intelligence, they are more malleable until later ages. Neuroscience shows that this malleability is associated with the slow development of the prefrontal cortex.[9] When attempted, adolescent remediation should focus on boosting character skills.

Skill development is a dynamic process. For example, boosting character skills early in

[9] Walsh (2005).

life increases the benefits of education later in life: More persistent students learn more. The levels of cognitive and character skills at any age depend on levels of those skills at younger ages and earlier investments.[10]

Inequality among families in parenting and lack of support given to children in schools are major contributors to inequality in skills. In light of this evidence, it is of great concern that so many American children are being born into disadvantage as measured by the quality of parenting and that the trend appears to be accelerating.[11] The traditional family with its secure environment for rearing children is under challenge. A major casualty of this trend is the quality of parenting available to disadvantaged children. It is unreasonable to expect schools to take on the huge burden of supplementing compromised family environments in addition to their traditional job of educating children.

Evidence from the social and biological sciences establishes the importance of the early years in fostering the skills that promote human flourishing. Families do much more than pass on their genes or put food on the table. Human development is a dynamic process of nature/nurture interactions that starts in the womb. Environments and skills interact to foster the development of later life skills and create who we are and what we become.

The foundations for adult success are laid down early in life. Many children raised in disadvantaged environments start behind and stay behind. Poverty has lasting effects on brain development, health, cognition, and character. Gaps in skills emerge early, before formal schooling begins. Waiting until kindergarten to address these gaps is too late. It creates achievement gaps for disadvantaged children that are costly to close.

Family disadvantage is poorly assessed by conventional measures of poverty that focus on family income flows and parental education. The absence of parental guidance, nourishment, and encouragement is the most damaging condition for child development. Absence of quality parenting—stimulation, attachment, encouragement, and support—is the true measure of child poverty.

[10] Cunha and Heckman (2007, 2008); Cunha, Heckman, and Schennach (2010).
[11] See McLanahan (2004) and Heckman (2008).

Children from disadvantaged homes are much more likely to drop out of high school. Single-parent families, compared to two-parent families, give less cognitive and emotional stimulation to their children (Moon, 2012), and as a consequence their children have diminished cognitive and character skills.[12] The evidence summarized in this paper suggests that skill deficits can be prevented by improving the early lives of disadvantaged children.

Yet, while important and often neglected in American public policy, the early years do not fully determine adult success. Children are resilient and quality parenting throughout childhood fosters the development and expression of skills. Schooling shapes both cognitive and character skills. Certain adolescent remediation programs appear to be effective.

A variety of adolescent interventions attempt to remediate early-life skill deficits. Alternatives to the traditional high school curriculum provide potential dropouts with training suitable to their interests and skills. Such programs can mold character, even if it is not their primary goal. Other interventions attempt to directly remediate character deficits after students drop out of school.

We aim to provide a sober account of what is known. We compare different skill enhancement strategies and consider the features that make some programs more successful than others. Unfortunately, the field of human development is marred by overzealous advocates who claim miracle fixes from their favored programs. They focus on one slice of the life cycle to the exclusion of others.

Slogans often replace hard evidence. Most evaluations of interventions have only short-term follow-ups. Many differ in their measures of outcomes and skills and target different demographic groups, so it is difficult to compare alternative programs. Despite these limitations, several studies suggest that interventions during the preschool years or in kindergarten improve character in a lasting way, some with annual rates of return (per annum yields) that are comparable to those from investment in the stock market in normal times.[13] Sev-

[12]See, for example, Carneiro and Heckman (2003), Cunha, Heckman, Lochner, and Masterov (2006), and Cunha and Heckman (2009).

[13]See Heckman, Moon, Pinto, Savelyev, and Yavitz (2010b).

eral adolescent interventions are promising, particularly those that combine education with work-related experience.[14]

This paper is organized as follows. Section 2 defines cognitive and character skills and discusses how they are measured. Section 3 summarizes the evidence on the predictive power of character skills. Section 4 presents an analysis of the skills needed in the workforce, as garnered from surveys of employers. Section 5 presents a framework for analyzing investment in skills over the life cycle. Sections 6– 12 review evidence about the efficacy of education, parental investment, and interventions in improving character skills from preschool, elementary school, and adolescence. Section 13 summarizes the paper.

2 Measures of Cognitive and Character Skills

2.1 Cognitive Skills

Measures of cognition have been developed and refined over the past century. Cognitive ability has multiple facets.[15] Psychologists distinguish between fluid intelligence (the rate at which people learn) and crystallized intelligence (acquired knowledge).[16] Achievement tests are designed to capture crystallized intelligence,[17] whereas IQ tests like Raven's progressive matrices (1962) are designed to capture fluid intelligence.[18,19]

This new understanding of cognition is not widely appreciated. Many use IQ tests, standardized achievement tests, and even grades as interchangeable measures of "cognitive

[14]See, for example, Kemple and Willner (2008) and Roder and Elliot (2011).

[15]See Carroll (1993) and Ackerman and Heggestad (1997) for discussions.

[16]See, for example, Nisbett, Aronson, Blair, Dickens, Flynn, Halpern, and Turkheimer (2012).

[17]Roberts, Goff, Anjoul, Kyllonen, Pallier, and Stankov (2000).

[18]Raven, Raven, and Court (1988). The high correlation between scores on intelligence tests and scores on achievement tests is in part due to the fact that both require intelligence and knowledge. Fluid intelligence promotes the acquisition of crystallized intelligence. Common developmental factors affect the formation of both skills.

[19]Carroll (1993) and Ackerman and Heggestad (1997) discuss more disaggregated facets of cognitive ability.

ability" or intelligence.[20] Scores on IQ tests and standardized achievement tests are strongly correlated with each other and with grades.[21] However, these general indicators of "cognition" measure different skills and capture different facets of cognitive ability.[22,23] In the following section we show that scores on these tests are also influenced by effort and character skills.

2.2 Measuring Character

Throughout this paper we use the term *character skills* to describe the personal attributes not thought to be measured by IQ tests or achievement tests. These attributes go by many names in the literature, including soft skills, personality traits, non-cognitive skills, non-cognitive abilities, character, and socio-emotional skills. These different names connote different properties.[24] "Traits" suggests a sense of permanence and possibly also of heritability. "Skills" suggests that these attributes can be learned. In reality, the extent to which these personal attributes can change lies on a spectrum. Both cognitive and character skills can change and be changed over the life cycle, but through different mechanisms and with different ease at different ages. We generally use the term *skill* throughout this paper because all attributes can be shaped.

Although character skills are overlooked in most contemporary policy discussions and in economic models of choice behavior, personality psychologists have studied these skills for the past century. Psychologists primarily measure character skills by using self-reported surveys or observer reports. They have arrived at a relatively well-accepted taxonomy of character skills called the Big Five, with the acronym OCEAN, which stands for: Openness

[20] This practice is true even among leading professional psychologists. For example, all of these measures are used as measures of intelligence in Flynn (2007), Nisbett (2009), and Nisbett et al. (2012).

[21] See Heckman and Kautz (2012).

[22] See Borghans, Golsteyn, Heckman, and Humphries (2011a).

[23] It is an irony of the testing literature that high school grades are more predictive of first-year college performance than SAT scores (Bowen, Chingos, and McPherson, 2009). The SAT and related tests were once thought to be a more objective measure of student quality than high school grades (Lemann, 1999).

[24] See Almlund et al. (2011) and Borghans et al. (2008) for comparisons of some of these different taxonomies.

to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Table 1 defines these skills and their multiple facets.[25] Some argue that the Big Five are the longitude and latitude of character skills, by which all more narrowly defined skills may be categorized.[26]

While the Big Five measures are now widely used in psychology, there are several other taxonomies, including the Big Three, the MPQ, and the Big Nine. They are conceptually and empirically related to the Big Five.[27] Other taxonomies, including psychopathology as measured by the DSM-IV and measures of temperament, have also been related to the Big Five.[28] Almlund et al. (2011) and Becker, Deckers, Dohmen, Falk, and Kosse (2012) summarize evidence showing that economic preference parameters are not closely related to the Big Five measures and apparently represent different skills that, along with the skills measured by psychologists, govern behavior.[29]

[25] See, for example, Borghans et al. (2008).

[26] Costa and McCrae (1992a).

[27] See Borghans et al. (2008) and Almlund et al. (2011) for comparisons of these taxonomies.

[28] See, for example, Cloninger, Svrakic, Bayon, and Przybeck (1999).

[29] A deeper issue, as yet not systematically investigated in the literature in economics or psychology, is whether the "traits" captured by the alternative measurement systems are the expression of a deeper set of preferences or goals. McAdams (2006) adds goals to the list of possible traits. Almlund et al. (2011) and Heckman and Kautz (2012) develop a model in which preferences and endowments of skills determine the effort applied to tasks. As shown in the next section, performance on tasks is the source of any measurement of a trait. Hence, in their framework, measures of traits are determined, in part, by preferences.

Table 1 The Big Five Domains and Their Facets

Big Five Personality Factor	American Psychology Association Dictionary Description	Facets (and correlated skill adjective)	Related Skills	Analogous Childhood Temperament Skills
Conscientiousness	"The tendency to be organized, responsible, and hardworking"	Competence (efficient), Order (organized), Dutifulness (not careless), Achievement striving (ambitious), Self-discipline (not lazy), and Deliberation (not impulsive)	Grit, Perseverance, Delay of gratification, Impulse control, Achievement striving, Ambition, and Work ethic	Attention/(lack of) distractibility, Effortful control, Impulse control/delay of gratification, Persistence, Activity*
Openness to Experience	"The tendency to be open to new aesthetic, cultural, or intellectual experiences"	Fantasy (imaginative), Aesthetic (artistic), Feelings (excitable), Actions (wide interests), Ideas (curious), and Values (unconventional)		Sensory sensitivity, Pleasure in low-intensity activities, Curiosity
Extraversion	"An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability"	Warmth (friendly), Gregariousness (sociable), Assertiveness (self-confident), Activity (energetic), Excitement seeking (adventurous), and Positive emotions (enthusiastic)		Surgency, Social dominance, Social vitality, Sensation seeking, Shyness*, Activity*, Positive emotionality, and Sociability/affiliation
Agreeableness	"The tendency to act in a cooperative, unselfish manner"	Trust (forgiving), Straight-forwardness (not demanding), Altruism (warm), Compliance (not stubborn), Modesty (not show-off), and Tender-mindedness (sympathetic)	Empathy, Perspective taking, Cooperation, and Competitiveness	Irritability*, Aggressiveness, and Willfulness
Neuroticism/ Emotional Stability	Emotional stability is "Predictability and consistency in emotional reactions, with absence of rapid mood changes." Neuroticism is "a chronic level of emotional instability and proneness to psychological distress"	Anxiety (worrying), Hostility (irritable), Depression (not contented), Self-consciousness (shy), Impulsiveness (moody), Vulnerability to stress (not self-confident)	Internal versus External, Locus of control, Core self-evaluation, Self-esteem, Self-efficacy, Optimism, and Axis I psychopathologies (mental disorders) including depression and anxiety disorders	Fearfulness/behavioral inhibition, Shyness*, Irritability*, Frustration, (Lack of) soothability, Sadness

Notes: *These temperament attritbutes may be related to two Big Five factors. Facets specified by the NEO-PI-R personality inventory (Costa and McCrae, 1992b). Adjectives in parentheses from the Adjective Check List (Gough and Heilbrun, 1983).
Source: Table adapted from John and Srivastava (1999).

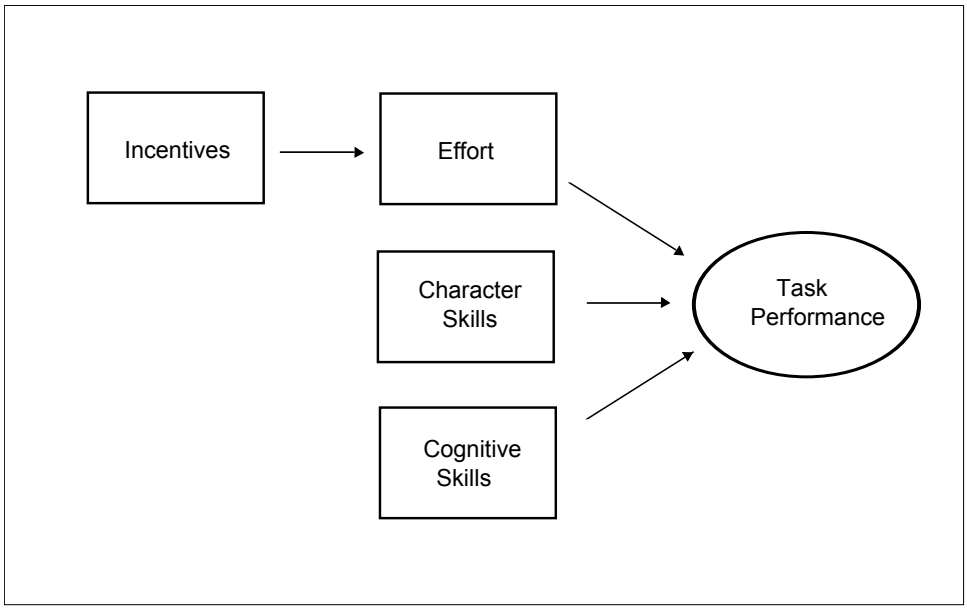
2.3 A Task-Based Framework for Identifying and Measuring Skills

One leading personality psychologist defines personality (character) traits (skills) as follows:

> *Personality traits are the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances.* (Roberts, 2009, 140)

Roberts' definition of personality ("character") suggests that all psychological measurements are calibrated on measured behavior or "tasks" broadly defined. A task could be an IQ test, a personality questionnaire, job performance, school attendance, completion of high school, participation in crime, or performance in an experiment. Figure 1 depicts how performance on a task can depend on incentives, effort, and cognitive and character skills. Performance on different tasks depend on these components to different degrees. People can compensate for their shortfalls in one dimension by having strengths in other dimensions.

Figure 1 Determinants of Task Performance



Many believe that personality skills can only be assessed by self-reported questionnaires that elicit skills like the Big Five. However, performance on any task or any observed behavior

can be used to measure skills.[30] For example, completing high school requires many other skills besides those measured by achievement tests, including showing up in school, paying attention, and behaving in class.[31]

Figure 1 suggests that inferring skills from performance on tasks requires standardization for all of the other contributing factors that produce the observed behaviors. The inability to parse and localize behaviors that depend on a single skill or ability gives rise to a fundamental problem of assessing the contribution of any skill to successful performance on a task. This problem is commonly ignored in empirical research that studies how psychological skills affect outcomes.[32]

There are two issues here. First, behavior depends on incentives created by situations. Different incentives elicit different amounts of effort on the tasks used to measure skills. Accurately measuring character skills requires standardizing for the effort applied in any task. Second, performance on most tasks depends on many skills. Not standardizing for incentives and other relevant skills can produce misleading estimates of any particular skill.

These problems are empirically important. For example, incentives partly determine scores on IQ tests. A series of studies conducted over the past 40 years show that incentives, like money or candy, can increase IQ scores, particularly among low-IQ individuals. The black–white gap in IQ can be completely eliminated by giving M&M candies for correct answers.[33] However, there is no evidence that this performance persists. It has not been shown that creating incentives for performance on one test improves performance on subsequent tests, and there is some evidence that it, in fact, may worsen subsequent performance (Deci and Ryan, 1985; Ryan and Deci, 2000). Duckworth, Quinn, Lynam, Loeber, and

[30]See Almlund et al. (2011).

[31]The idea of using behaviors to measure character is old. Ralph Tyler suggested using measures of behavior to capture character skills in his first proposal for the National Assessment of Educational Progress tests. See Tyler (1973) and Rothstein, Jacobsen, and Wilder (2008). This idea is being been pursued in the recent literature (Heckman, Humphries, Urzúa, and Veramendi, 2011; Jackson, 2013).

[32]See Borghans et al. (2011a), Almlund et al. (2011), and Heckman and Kautz (2012) for discussions of this problem.

[33]See Ayllon and Kelly (1972); Borghans, Meijers, and ter Weel (2008); Breuning and Zella (1978); Clingman and Fowler (1976); Edlund (1972); Holt and Hobbs (1979); Larson, Saccuzzo, and Brown (1994); Segal (2008). This evidence is summarized in Borghans et al. (2008) and Almlund et al. (2011).

Stouthamer-Loeber (2011) have supplemented these studies by showing that observer-rated motivation predicts IQ scores.

Not all persons respond with equal strength to incentives. Research by Borghans et al. (2008) and Segal (2008) shows that the responsiveness of persons to incentives on IQ and achievement tests depends on their character skills.

The recent literature shows that character skills predict standardized achievement test scores, which many analysts assume are good measures of intelligence.[34] Figures 2 and 3 show how the variance in the scores on two achievement tests, the Armed Forces Qualifying Test (AFQT) and the closely related Differential Aptitudes Test (DAT),[35] can be decomposed into IQ and character measures. Character skills explain a substantial portion of the variability in both AFQT scores and DAT scores.[36] Character skills explain the variance in achievement scores above and beyond the variance that IQ explains when both character and IQ are included in a regression. These findings caution the interpretation that standardized achievement tests only measure cognitive ability. They also capture character skills.[37] Ironically, *The Bell Curve* by Herrnstein and Murray, which uses achievement tests as a measure of intelligence, implicitly shows the power of both character and cognition in shaping life outcomes.

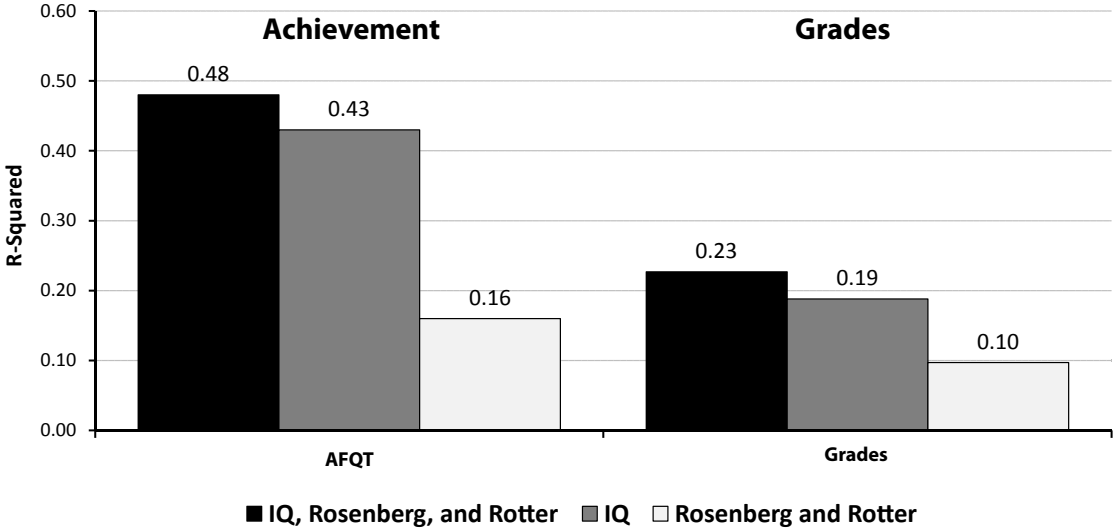
[34]See, for example, Nisbett (2009).

[35]The correlation between DAT and AFQT scores in the National Longitudinal Study of Youth 1979 (NLSY79) is 0.75 (Borghans, Golsteyn, Heckman, and Humphries, 2011b). Friedman and Streicher (1985) estimate correlations between 0.65 and 0.82 in a sample of high school sophomores and juniors. Kettner (1976) estimates correlations between DAT and the AFQT subtests of 0.76 to 0.89 in a sample of juniors and seniors.

[36]The lower explained variance in the sample with DAT is likely a consequence of restriction on range. The DAT data come from a single school, whereas the AFQT data come from a national sample.

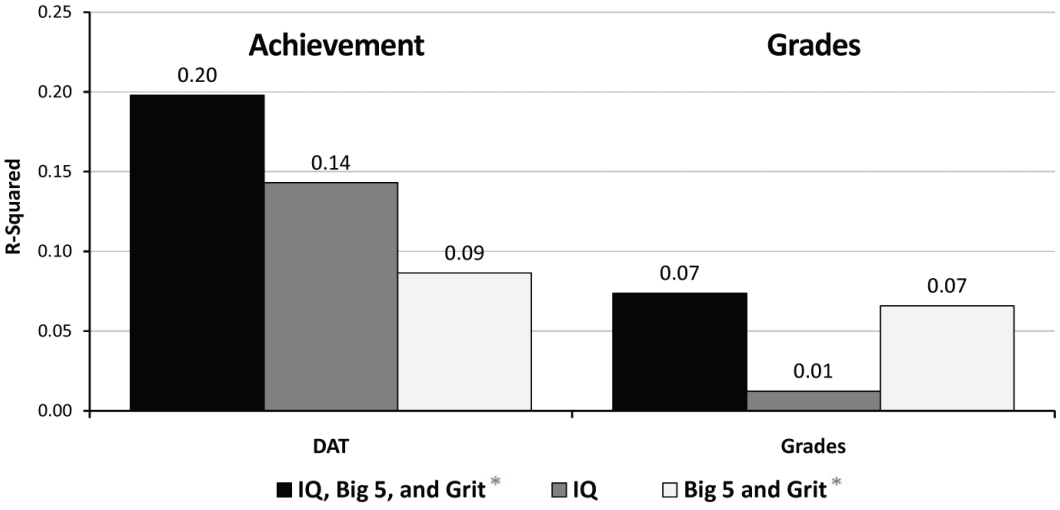
[37]In the Stella Maris data, Openness to Experience is strongly correlated with IQ. See Borghans et al. (2011b).

Figure 2 Decomposing Variance Explained for Achievement Tests and Grades into IQ and Character [NLSY79]



Source: Borghans et al. (2011a).
Notes: Personality is measured by the Rotter and Rosenberg personality inventories. These are only a subset of the personality inventories described in Table 1. Locus of Control is based on the four-item abbreviated version of the Rotter Internal- External Locus of Control Scale. This scale is designed to measure the extent to which individuals believe that they have control over their lives through self-motivation or self-determination (internal control) as opposed to the extent to which individuals believe that the environment controls their lives (external control). The self-esteem scale is based on the 10-item Rosenberg Self-Esteem Scale. This scale describes a degree of approval or disapproval toward oneself. Rotter was administered in 1979. The ASVAB and Rosenberg were administered in 1980. AFQT is constructed from the Arithmetic Reasoning, Word Knowledge, Mathematical Knowledge, and Paragraph Comprehension ASVAB subtests. IQ and GPA are from high school transcript data. AFQT, Rosenberg, and Rotter have been adjusted for schooling at the time of the test conditional on final schooling, as laid out in Hansen, Heckman, and Mullen (2004). IQ is pooled across several IQ tests using IQ percentiles. GPA is the individual's core subject GPA from 9th grade. Sample excludes the military oversample. R-squared is a measure of the proportion of variance in the outcome explained by the designated variable.

Figure 3 Decomposing Variance Explained for Achievement Tests and Grades into IQ and Character: Stella Maris Secondary School, Maastricht, Holland



Source: Borghans et al. (2011a).
Note: Grit is a measure of persistence on tasks (Duckworth, Peterson, Matthews, and Kelly, 2007).

2.4 Reference Bias

Answers from self-reports can be misleading when comparing levels of personality skills across different groups of people. Most personality assessments do not anchor their measurements in any objective outcome.[38] For example, the German Socio-Economic Panel (GSOEP) survey asks respondents to rate themselves on the following statement: *"I see myself as someone who tends to be lazy"* (Lang, John, Lüdtke, Schupp, and Wagner, 2011). The scale ranges from 1 = "strongly disagree" to 7 = "strongly agree." In answering this question, people must interpret the definition of *"lazy,"* which likely involves comparing themselves to other people. If different groups have different standards or reference points, comparing traits across groups can be highly misleading. Laziness may mean different things to different groups of people.

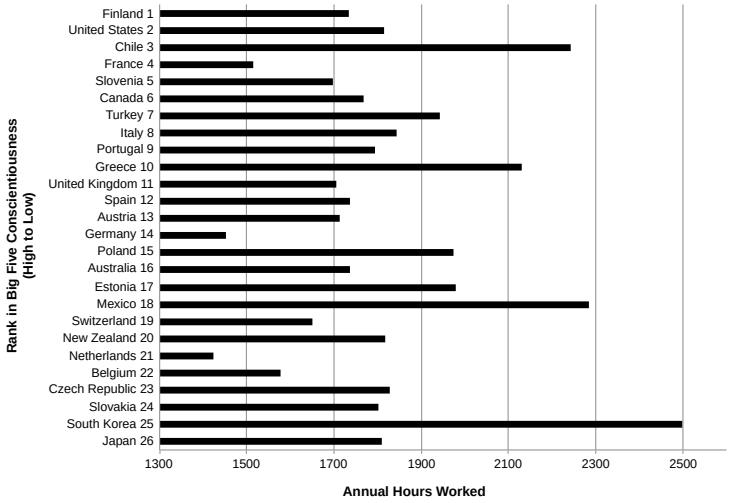
This measurement problem—sometimes called reference bias—is empirically relevant.[39] Schmitt, Allik, McCrae, and Benet-Martínez (2007) administer a Big Five personality questionnaire to groups of people in a variety of different countries. Using their estimates, Figure 4 shows how Organization of Economic Cooperation of Development (OECD) countries rank in Conscientiousness (from high to low). The bars display the average number of hours that people work in the country. The results are surprising. South Korea ranks second to last in terms of Conscientiousness but also ranks first in the number of hours worked. South Korea is not an anomaly. Country-level reports of Big Five Conscientiousness are unrelated to the number of hours worked. The rank correlation between hours worked and Conscientiousness across countries is negative, though statistically insignificant.[40]

[38] These are called Likert scales (Likert, 1932).

[39] Reference bias is also problematic in health surveys that use self-reported, subjective health assessments. See Groot (2000).

[40] $r = -0.07$ ($p = 0.73$).

Figure 4 National Rank in Big Five Conscientiousness and Average Annual Hours Worked



Source: The Conscientiousness ranks come from Schmitt et al. (2007). These measures were taken in 2001 (Schmitt, 2002). The hours worked estimates come from Organisation of Economic Cooperation and Development (2001).
Note: Several countries are omitted due to lack of data.

These findings are relevant for constructing measurements of character skills. Achievement tests miss important skills, suggesting that students, teachers, and schools should not be evaluated solely by achievement tests. Would the self-reported Big Five be a useful supplement to school evaluations? The possibility of reference bias suggests that it might not. Self-reports of Conscientiousness might measure different things for different schools.

Psychologists have attempted to address the problem of reference bias.[41] Some surveys include vignette-based questions that attempt to standardize for aspects of the culture or situation. They attempt to frame questions so that the people in the survey answer within a common situation. However, this approach might not work well for evaluating schools, especially if teachers have incentives to coach children on taking these tests so that they score better and give answers perceived to be positive. Direct use of standard psychological

[41]For further discussion of reference bias, see Duckworth (2012) and her references: Goldammer (2010); Heine, Buchtel, and Norenzayan (2008); Heine, Kitayama, Lehman, Takata, Ide, Leung, and Matsumoto (2001); Heine, Lehman, Peng, and Greenholtz (2002); Naumann and John (2011); Peng, Nisbett, and Wong (1997); and Schmitt et al. (2007).

measures can be problematic.[42]

2.5 Measuring Skills Using Behaviors

Ralph Tyler, a pioneer of achievement testing, recognized its limitations. He suggested using measures of behavior such as performance, participation in student activities, and other observations by teachers and school administrators to complement achievement tests when evaluating students and schools. Several recent papers show that this is a promising approach. Heckman, Pinto, and Savelyev (2013) show that teacher ratings of elementary schoolchildren's behaviors are strong predictors of adult outcomes and that early childhood interventions promote the character skills measured by these ratings. Heckman et al. (2011) estimate the causal effect of cognitive and socio-emotional (character) skills on a variety of outcomes. They measure socio-emotional (character) skills using risky and reckless behaviors measured in the adolescent years.[43] Motivated by findings such as those reported in Figure 2 and Figure 3, they develop and apply methods to use high school grades to measure both cognition and character. They show that character skills promote educational attainment, beneficial labor market outcomes, and health.

Jackson (2013) studies the effect of teachers on student cognition and character. In a fashion similar to Heckman et al. (2011), Jackson measures cognitive skills using achievement test scores, while measuring noncognitive (character) skills using absences, suspensions, grades, and grade progression. These measures of character predict adult outcomes with a strength similar to measures of cognitive ability. His measures of character are commonly available from the administrative records of schools.

Similar to Ralph Tyler's suggestion to use participation in extracurricular activities,

[42]In an attempt to address reference bias, some psychologists measure skills using behaviors. Heine et al. (2008) examine cross-country differences in Conscientiousness using objective measures, including walking speed, postal workers' speed, and the accuracy of clocks in public banks.To measure walking speed, researchers timed how long it took for a random sample of people to walk 60 feet in public areas. Postal workers' speed was assessed by measuring how long it took for postal workers to sell stamps.

[43]The measure of risky and reckless behavior is based on whether adolescents engaged in any of the following behaviors: stealing from a store, purposefully damaging property, taking something worth less than $50, or conning someone.

Lleras (2008) measures noncognitive (character) skills in part by using tenth grade participation in sports, academic clubs, and fine arts activities. Participation in these activities predicts educational attainment 10 years later, even after controlling for cognitive ability as measured by achievement tests.

Criminologists have also debated about whether it is better to measure self-control with self-reported psychological scales or with objective behaviors. The publication of Gottfredson and Hirschi's *A General Theory of Crime* (1990) launched a series of studies on the link between self-control and crime. In that study, they posit that a single factor, self-control, predicts much of the variance in criminal outcomes.

There is a divide in this literature. Subsequent studies have measured self-control using psychological scales, while others have used behavioral measures. A meta-analysis by Pratt and Cullen (2000) finds that behavioral measures are at least as good at predicting crime as are measures based on self-reported taxonomies. In a similar vein, Benda (2005) uses both types of measures in the same study and finds that behavioral measures predict crime better than psychological scales.

Hirschi and Gottfredson (1993) suggest that objective behavioral measures might be preferred to self-reports, partly because the act of filling out a survey requires some level of self-control. Answering survey questions is another task that relies on skills beyond the ones targeted by the survey.

Some criticize this approach and argue that it is tautological to use measures of behavior to predict other behavior even though the measures are taken early in life to predict later life behaviors.[44] As suggested by Figure 1, all tasks or behaviors can be used to infer a skill as long as the measurement accounts for other skills and aspects of the situation. Self-reported scales should not be assumed to be more reliable. The question is which measurements are most predictive and can be implemented in practice. The literature suggests that there are objective measurements of character that are not plagued by reference bias.

[44]See the discussion in Pratt and Cullen (2000) and Benda (2005).

2.6 Are Character Skills Stable?

Many have questioned whether there are stable character skills. The publication of Walter Mischel's 1968 book, *Personality and Assessment*, gave rise to a heated "personality–situation" debate within psychology, which pitted social psychologists who favored situational factors as primary determinants of behavior against personality psychologists who considered stable personality (character) traits (skills) as more consequential. Mischel argued that aspects of situations overshadow any effect of personality (character) traits (skills) on behavior.[45] Ironically, Mischel himself later demonstrated the stability of character skills over the life cycle in his celebrated "marshmallow experiment."[46]

A large body of evidence reviewed in Almlund et al. (2011) shows that stable character skills exist and are predictive of many behaviors.[47] An early paper by Epstein (1979) presents compelling evidence that, averaging over tasks and situations, people act in a predictable fashion with a high level of reliability of average behavior ("measured character") across situations.[48]

In addition, measures of character skills tend to be about 30%–60% heritable, suggesting that something tied to the person, rather than the situation, influences behavior (Bouchard and Loehlin, 2001).[49] Evidence from neuroscience suggests that expression of different skills is linked to different regions of the brain (see Canli, 2006, and DeYoung, Hirsh, Shane, Papademetris, Rajeevan, and Gray, 2010).

[45] This point is echoed in behavioral economics. See Thaler (2008).

[46] A participant (usually a child) was given a marshmallow. The experimenter left the room and told the participant that he or she would receive a second marshmallow if he or she resisted consuming the marshmallow until the experimenter returns. The length of time that the participant waits is a measure of short-term discounting. The children who could wait had much better lifetime outcomes. (For a recent discussion of this study, see Mischel, Ayduk, Berman, Casey, Gotlib, Jonides, Kross, Teslovich, Wilson, Zayas, and Shoda, 2011.)

[47] See the special issue of *Journal of Research in Personality* (43) entitled "*Personality and Assessment at Age 40*" for a recent discussion.

[48] R^2 of 0.6–0.8, where R^2 is a measure of variance explained.

[49] Devlin, Daniels, and Roeder (1997) suggest that traditional estimates of the heritability of IQ may be inflated because they fail to take into account the effect of the environment on conditions in the maternal womb. See also Rutter (2006) and an emerging literature on epigenetics (Jablonka and Raz, 2009).

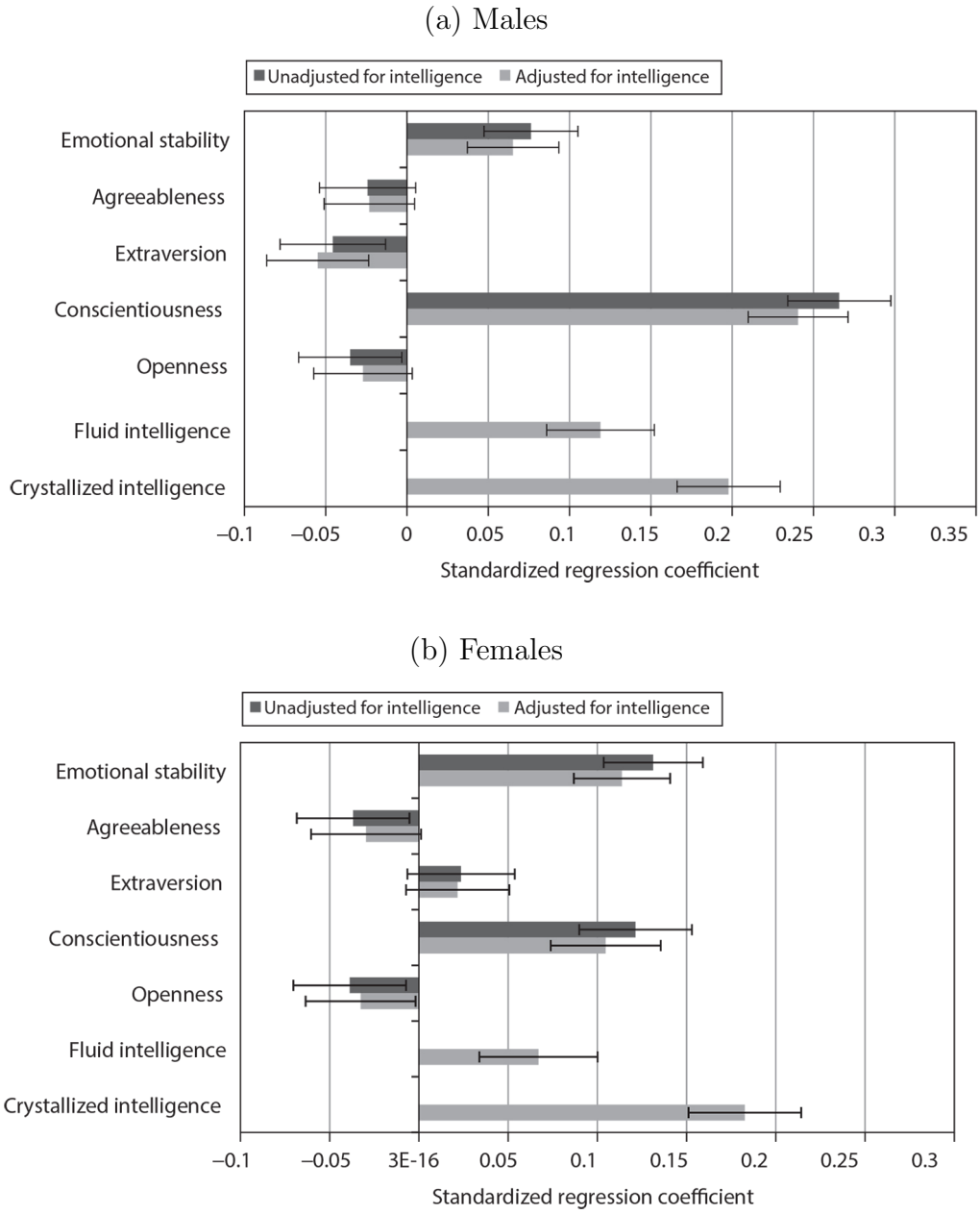
3 Predictive Power of Character

A substantial body of evidence shows that character skills predict a wide range of life outcomes, including educational achievement, labor market outcomes, health, and criminality. For many outcomes, the predictive power of character skills rivals that of measures of cognitive ability. Of the Big Five, Conscientiousness—the tendency to be organized, responsible, and hardworking—is the most widely predictive across a variety of outcomes.[50]

Given that school requires hard work and perseverance, it is not surprising that Conscientiousness is associated with educational attainment. Figure 5 shows standardized coefficients from a regression of the Big Five on educational attainment from the German Socioeconomic Panel (GSOEP). The darker lines show effects without adjusting for measures of intelligence, and the lighter lines show effects after additionally adjusting for measures of fluid and crystallized intelligence. Conscientiousness predicts years of schooling with the same strength as the measures of intelligence.

[50]See Almlund et al. (2011); Borghans et al. (2008); Heckman and Kautz (2012); Roberts et al. (2007).

Figure 5 Association of the Big Five and Intelligence with Years of Schooling in GSOEP

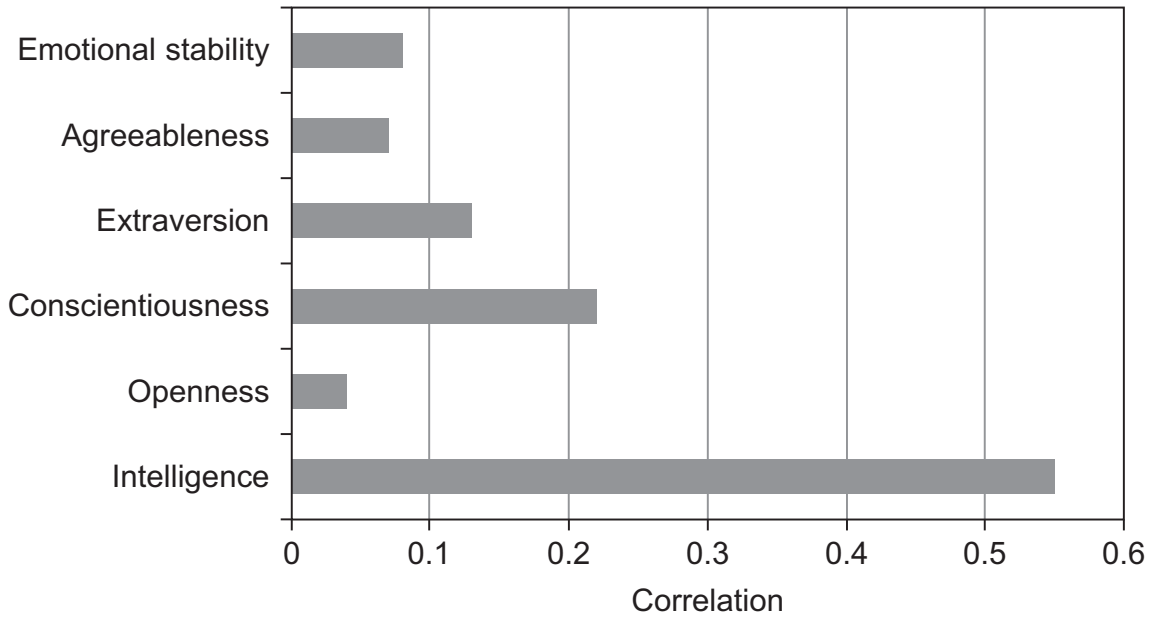


Source: German Socio-Economic Panel (GSOEP), waves 2004–2008, Almlund et al. (2011).
Notes: The figure displays standardized regression coefficients from a multivariate regression of years of school attended on the Big Five and intelligence, controlling for age and age squared. The bars represent standard errors. The Big Five coefficients are corrected for attenuation bias. The Big Five were measured in 2005. Years of schooling were measured in 2008. Intelligence was measured in 2006. The measures of intelligence were based on components of the Wechsler Adult Intelligence Scale (WAIS). The data is a representative sample of German adults between the ages of 21 and 94.

Aspects of job performance are also related to academic performance. Both require com-

pleting work on a schedule and involve intelligence to varying degrees. As with academic performance, numerous studies and meta-analyses have found that Conscientiousness is associated with job performance and wages (Barrick and Mount, 1991; Hogan and Holland, 2003; Nyhus and Pons, 2005; Salgado, 1997). Figure 6 presents correlations of the Big Five and IQ with job performance. Of the Big Five factors, Conscientiousness is the most strongly associated with job performance but is about half as predictive as IQ. Conscientiousness, however, may play a more ubiquitous role than IQ. The importance of IQ increases with job complexity (the information processing requirements of the job). Cognitive skills are more important for professors, scientists, and senior managers than for semiskilled or unskilled laborers (Schmidt and Hunter, 2004). The importance of Conscientiousness does not vary much with job complexity (Barrick and Mount, 1991); this suggests that it applies to a wider spectrum of jobs.

Figure 6 Associations with Job Performance



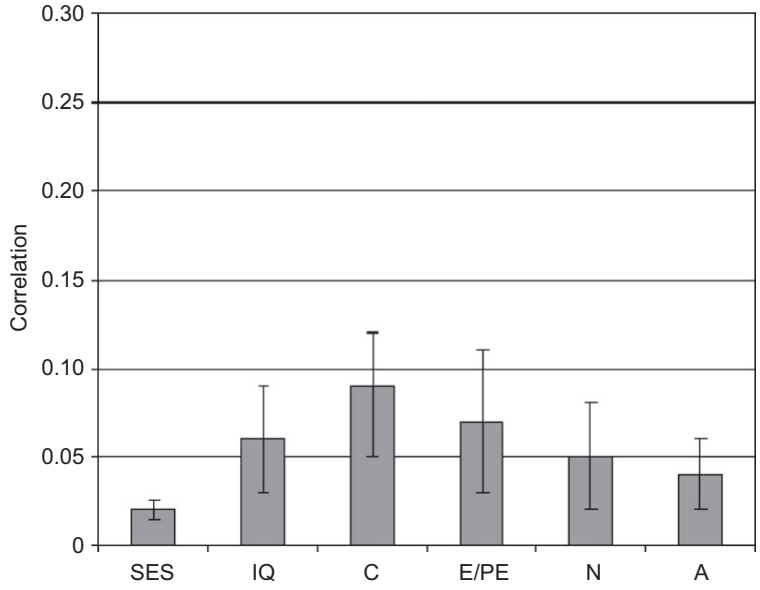
Source: The correlations reported for personality traits come from a meta-analysis conducted by Barrick and Mount (1991). The correlation reported for intelligence comes from Schmidt and Hunter (2004).
Notes: The values for personality are correlations that were corrected for sampling error, censoring, and measurement error. Job performance was based on performance ratings, productivity data, and training proficiency. The authors do report the timing of the measurements of personality relative to job performance. Of the Big Five, the coefficient on Conscientiousness is the only one statistically significant, with a lower bound on the 90% credibility value of 0.10. The value for intelligence is a raw correlation.

Measures of character skills rival IQ and measures of socioeconomic status in predicting longevity.[51] Roberts et al. (2007) review evidence from 34 different studies on the predictive validity of the Big Five personality measures, relative to that of cognitive ability and socioeconomic status. Most studies in their meta-analysis control for relevant background factors, including gender and severity of disease. Roberts and colleagues convert the results of each study into correlation coefficients that can be compared across studies. Figure 7 presents results from their analyses. Conscientiousness is a stronger predictor of longevity than any other Big Five trait and a stronger predictor than either IQ or socioeconomic status.[52] In general, skills related to Conscientiousness, Openness to Experience, and Agreeableness are associated with longer lives, whereas those related to Neuroticism are associated with shorter

[51]For a recent study, see Friedman and Martin (2011).
[52]The timing of the measurements of character relative to the outcomes varies by study.

lives.[53] The magnitudes of the relationships vary across studies.

Figure 7 Correlations of Mortality with Character, IQ, and Socioeconomic Status (SES)

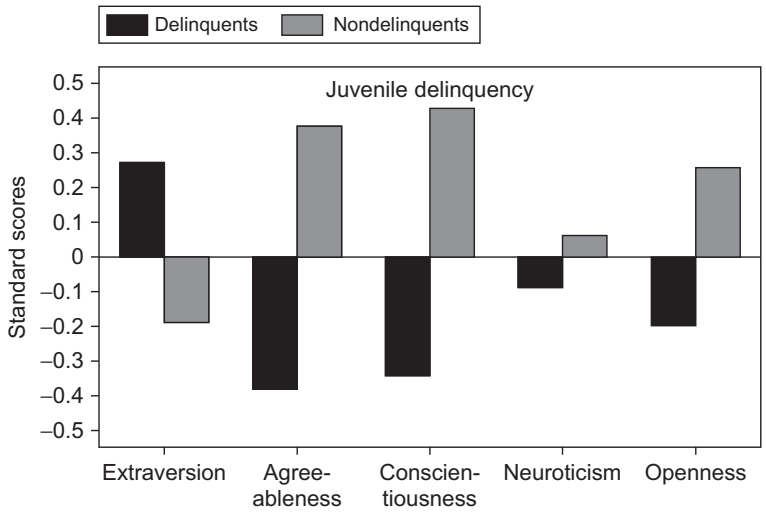


Source: Roberts et al. (2007).
Notes: The figure represents results from a meta-analysis of 34 studies. Average effects (in the correlation metric) of low socioeconomic status (SES), low IQ, low Conscientiousness (C), low Extraversion/Positive Emotion (E/PE), Neuroticism (N), and low Agreeableness (A) on mortality. Error bars represent standard errors. The lengths of the studies represented vary from 1 year to 71 years.

Of the Big Five, Conscientiousness and Agreeableness are most predictive of criminality. Figure 8 illustrates that in a sample of at-risk youth, boys who committed severe delinquent behaviors were more than three quarters of a standard deviation lower in Agreeableness and Conscientiousness, as measured by mothers' reports at age 12 or 13, than boys who had minor or no delinquent behaviors up to that age (John, Caspi, Robins, and Moffitt, 1994).

[53]See Boyle, Williams, Mark, Brummett, Siegler, and Barefoot (2005); Friedman and Martin (2011); Kern and Friedman (2008); Kubzansky, Sparrow, Vokonas, and Kawachi (2001); Martin, Friedman, and Schwartz (2007); Mroczek and Spiro (2007); Schulz, Bookwala, Knapp, Scheier, and Williamson (1996) .

Figure 8 Juvenile Delinquency and the Big Five



Source: John et al. (1994).
Notes: Delinquents are those who have committed at least one of the following: breaking and entering, strong-arming, or selling drugs. Nondelinquents have committed no more than the following crimes: stealing at home, vandalism at home, or theft of something less than $5. The y-axis reports mean differences in standardized scores of the Big Five measures based on mothers' reports. The measures were taken at ages 12–13 and reflect cumulative delinquent behavior.

As with most studies in personality psychology, the evidence presented in Figures 5– 8 does not address the question of causality; that is, do measured skills *cause* (rather than just predict) outcomes? Empirical associations are not a reliable basis for policy analysis. As previously noted and summarized in Figure 1, multiple skills, effort, and acquired skills generate performance in a given task. Many studies in personality psychology do not control for all of the factors that produce performance on measured tasks. They equate measures of outcomes with the skill being measured.[54] This practice can lead to a substantial bias in inference about the importance of any particular skill. Our survey of the intervention literature in the next sections presents evidence on the causal relationship between skills and outcomes.

[54]Selecting measures and verifying them is part of the mysterious and inherently subjective process of "construct validity" in psychology. For a discussion, see Borghans et al. (2008)

4 The Skills Needed for Success in the Labor Market[55]

Another perspective on the importance of character skills comes from surveys of employers and workers. In 1991, the Secretary's Commission on Achieving Necessary Skills (SCANS) conducted an extensive analysis of which skills workers needed in the American workforce.[56] The Commission researched the literature, consulted with experts, and conducted detailed interviews with workers and/or supervisors in 50 occupations. The interviews rated the importance of various skills in the context of illustrative tasks and tools on the job. Using these sources, the Commission categorized necessary skills into basic skills, thinking skills, personal qualities, and a set of workplace competencies. In addition to reading, writing, and math skills, basic skills include listening and speaking. The thinking skills cover creative thinking, decision making, problem solving, reasoning, and the ability to learn. SCANS specifies that personal qualities include responsibility, self-esteem, sociability, self-management, integrity, and honesty. SCANS identifies five groups of workplace competencies: the ability to allocate resources (time, money, facilities), interpersonal skills (such as teamwork, teaching others, leadership), the ability to acquire and to use information, the ability to understand systems, and the ability to work well with technology.

Employer surveys reinforce the importance of skills that go well beyond academic skills. In a survey of 3,200 employers in four large metropolitan areas conducted in the 1990s, employers reported that such personal qualities as responsibility, integrity, and self-management are as important as or more important than basic skills (Holzer, 1997). In another employer survey undertaken in the mid-1990s of 3,300 businesses (the National Employer Survey), employers ranked attitude, communication skills, previous work experience, employer recommendations, and industry-based credentials above years of schooling, grades, and test scores as part of the skills needed for success in the workplace (Zemsky, 1997).

Nonacademic skills are especially critical for entry level and hourly workers. Of employers

[55]We have benefitted from the commentary of Robert Lerman in preparing this section. See Lerman (2013).

[56]Secretary's Commission on Achieving Necessary Skills (1992).

drawn from a national sample in the United States in 1996, 69% reported rejecting hourly applicants because they lacked basic employability skills, such as showing up every day, coming to work on time, and having a strong work ethic. This percentage is more than double the percentage of rejecting applicants due to inadequate reading and writing skills. Rejections for not passing a drug test were almost as common as rejections for lack of literacy skills.[57] In a 2007 survey of employers in Washington State, about 60% reported difficulty in hiring. They experienced less difficulty finding workers with adequate reading, writing, and math skills than with appropriate occupational, problem solving, teamwork, communication, and adaptability skills as well as positive work habits and a willingness to accept supervision.[58]

Evidence from the United Kingdom supports these findings. A 1998 survey of 4,000 employers found that the four skills found most lacking in 16 to 24-year-olds were technical and practical skills, general communication skills, customer handling skills, and teamwork skills.[59] At the bottom of the list were numeracy and literacy skills. In a 2002 survey of 4,000 employers in the UK, 23% of employers reported a significant number of their staff were less than fully proficient at their jobs. Skill shortfalls were most common in communication, teamwork, other technical and practical skills, customer handling, and problem solving and least common in numeracy and literacy.[60]

Consistent with these findings, the Confederation of British Industry defines employability as (1) values and attitudes compatible with the work, including a desire to learn, to apply that learning, to improve, and to take advantage of change; (2) basic skills (literacy and numeracy); (3) key skills (communication, application of numbers, information technology, improving one's own learning and performance, working with others, problem solving) sufficient for the needs of the work; (4) other generic skills such as modern language and customer service skills; and (5) job-specific skills and the ability to manage one's own career.

[57] Barton (2006).
[58] Washington Workforce Training Board (2008).
[59] Westwood (2004).
[60] Hillage, Regan, Dickson, and McLoughlin (2002).

An ethnographic approach provides some revealing examples of how skills are used in context and how nonacademic skills are often developed and used as part of a "community of practice".[61] In addition to formal knowledge, Nelsen (1997) points out that workplaces require not only formal knowledge, facts, principles, theories, and math and writing skills, but also informal knowledge embodied in heuristics, work styles, and contextualized understanding of tools and techniques. In her revealing case study of auto repair workers, Nelsen argues that social skills of new workers are very important for learning the informal knowledge of experienced workers, as captured in stories, advice, and guided practice.

5 A Framework for Understanding Interventions

A variety of strategies have been advocated for promoting character and cognitive skills. Different advocates press their favorite strategies. Policymakers have limited resources, so it is important to identify which programs are most effective. The rate of return—the per-period yield of an investment—provides a convenient and interpretable summary of the efficacy of competing programs. An alternative to funding an early education program might be to invest the resources in the stock market and to use the resources plus the accrued interest to fund a program for adolescents.

In this section, we present a conceptual scheme for understanding the dynamics of skill development.[62] Figure 9 illustrates the basic idea. In this framework, cognitive and character skills are produced by investments at different stages of the life cycle. Skills at birth depend on prenatal investments and inherited traits. Skill formation at later ages depends on the stock of skills acquired earlier as well as prior investments. For example, a child who has a better attention span more easily acquires more cognitive skills from classroom instruction. This concept is called *self-productivity* and is summarized by the motto "skills beget skills."

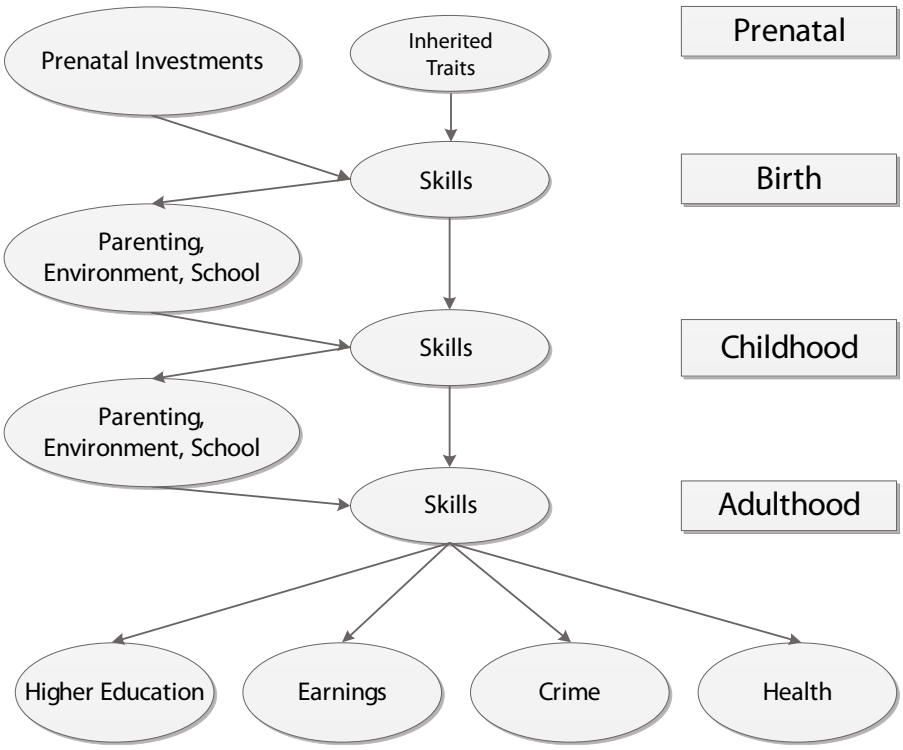
Investments (parenting, environment, and schools) also affect skills. The efficacy of in-

[61] Stasz (2001).

[62] These ideas are formalized in Cunha and Heckman (2007, 2008); Cunha et al. (2010).

vestment is moderated by the stock of skills at any age. The benefits of an investment depend on the current level of skills—a phenomenon that economists call *static complementarity*. More motivated children learn more. In addition, investment today increases the stock of future skills, which in turn increases the return to future investments. Economists call this phenomenon *dynamic complementarity*. This channel increases the returns to early investments because it makes future investments more productive. For this reason, Cunha et al. (2010) show that it is economically efficient to invest in the most disadvantaged young children because it raises their payoffs from future investments. See Heckman and Mosso (2013) for a more complete discussion of static and dynamic complementarity.

Figure 9 Framework for Understanding Skill Development



This framework recognizes that different skills might be relatively easy to shape at different stages of the life cycle. *Sensitive periods* for a given skill are periods when investments are relatively more productive. *Critical periods* for a particular skill are periods when investment during any other period is not productive.

Figure 9 illustrates why understanding the effects attributable to specific interventions is such a challenging task. Most of the empirical studies that we review only investigate the interventions received in a slice of the life cycle. They do not connect the links or correct for the effects of later investment in producing the outcomes attributed to early investments.

6 Summary of Empirical Evidence on the Efficacy of Interventions

This section summarizes the empirical evidence from a variety of interventions ranging from targeting prenatal infants to targeting young adults. In the subsequent sections, we discuss some of these programs in greater detail. In these discussions, we focus on programs that have been well studied, have long-term follow-ups, have been widely adopted, or offer unique insights. Descriptions and discussions of other programs can be found in the Web Appendix.[63]

For three reasons, evaluating and comparing the evidence from intervention programs is challenging. First, many interventions are only evaluated with short-term follow-ups, which could lead to upward-biased estimates of returns if the benefits eventually dissipate or to downward-biased estimates of the returns if the effects of the programs appear later in life. Second, not all studies measure the same outcomes. Ideally, all studies would report outcomes in terms of the rate of return of the program. Reported outcomes often differ across studies. Many studies only consider the effect of an intervention on a few outcomes. Without knowing the range of outcomes affected, it is difficult to calculate a rate of return.

[63]See `https://heckman.uchicago.edu/fostering_skills_appendix`.

Third, many programs target specific demographic groups. Applying the findings from one group to another might be problematic if groups differentially benefit from programs.

Table 2 summarizes the effects of each intervention discussed in this section. The table displays information about the nature of the intervention, the quality of the evaluation, the effects on later life outcomes, and estimates of the rate of return and cost-benefit ratio when available. The squares in the "Components" columns indicate the extent to which the program and the evaluation of it have the features defined in the table. The dots in the "Effects on Outcomes" columns indicate the extent to which the program influenced skills and outcomes. (The notes at the bottom of the table define the symbols and abbreviations used.)

Three striking patterns emerge about the nature of the programs and the quality of the available evaluations of them. First, as a group, early childhood and elementary school programs have longer follow-ups. All of the early childhood or elementary school programs in Table 2 have evaluations that follow participants for at least 10 years and many follow them more than 20 years, whereas only two evaluations of adolescent programs follow participants for at least 10 years (the longest is 12).

Second, early childhood programs tend to measure cognitive and character skills in addition to a variety of later-life outcomes, whereas many of the adolescent evaluations focus solely on labor market outcomes. Because of these features of data availability, we can better understand the sources of the effects on adult outcomes of early childhood programs by considering how these interventions produce skills. Due to the absence of measures of skills for many adolescent interventions, understanding these programs requires examining the curricula of the programs themselves, for example, whether the program seeks to foster cognitive or character skills.

Third, selection into programs differs by the age of intervention. In most early childhood evaluations, the programs first contact parents to participate and then parents opt into the program. In contrast, in most adolescent evaluations, participants themselves chose to enter

the program.

Table 2 also suggests certain features of effective programs. Only very early interventions (before age 3) improve IQ in a lasting way, consistent with the evidence that early childhood is a critical period for cognitive development (see Knudsen, Heckman, Cameron, and Shonkoff, 2006). The most successful interventions target preschoolers and primary school children. They improve later-life outcomes by developing character skills.

Programs that target adolescents have not been established to be as effective as programs that target earlier ages, in part because there have been fewer long-term evaluations of them. Several of the successful adolescent mentoring or residential programs improve labor market and social outcomes, but have relatively short follow-ups. The two programs with the longest follow-ups improve outcomes in the short run, but the benefits fade after a few years. These programs alter participants' environments and incentives during the intervention, which could influence their behavior in the short term without having a lasting effect.

The most promising adolescent programs integrate aspects of work into traditional education. Such programs break down the rigid separation between school and work that characterizes the American high school.

High schools create an adolescent society with values distinct from those of the larger society and removed from the workplace.[64] Even in affluent communities, the adolescent society has an anti-academic, anti-achievement bias. It was not until 1940 that more than half of each birth cohort graduated from high school.[65] In earlier times, adolescents took apprenticeships and jobs where they were supervised and mentored by adults. Mentoring involved teaching valuable character skills—showing up for work, cooperating with others, and persevering on tasks. These skills could be fostered in high schools, but with the relaxation of discipline in the schools, it is more difficult to do so.[66]

The apparent success of apprenticeship programs might arise in part from their cultivation

[64]See Coleman (1961).
[65]See Goldin and Katz (2008).
[66]See Arum (2005).

of character skills. The attachment of a supervisor to an apprentice helps create character in a version of the attachment bond between parent and child.[67]

7 Early-Life Interventions that Begin Before Formal Schooling

This section discusses the most influential interventions listed in Table 2 that start before children enroll in kindergarten. We divide the early interventions into two categories: (1) infant programs and model preschools that target infants by directly providing cognitive and socio-emotional stimulation to children, by instructing young mothers or by providing center-based care to a specific population and (2) large-scale programs that have been implemented in schools and are relatively less expensive than center-based care. Many early interventions have successfully boosted cognitive skills, character skills, or both.

Given that there are more differences among evaluations than there are evaluations, it is difficult to understand exactly *why* some programs are more successful than others. Nevertheless, certain patterns emerge. Only programs that start before the age of 3 have had lasting impacts on IQ. Some of the most successful programs have had no long-term impact on IQ but have improved later-life outcomes by improving character skills. As a general rule, large-scale programs have been less successful than smaller-scale programs, but the large-scale programs differ from small-scale programs in important ways.

7.1 Infant Programs and Model Preschools

7.1.1 Nurse-Family Partnership

Evidence from the Nurse-Family Partnership (NFP) program shows how improving prenatal conditions and early parenting can improve child skills and outcomes. NFP aims at first-

[67] See Bowlby (1951); Sroufe (1997); Sroufe, Egeland, Carlson, and Collins (2005).

Table 2 Summary of Effects for Main Interventions

	Participant/Evaluation Characteristics						Components					Effects on Outcomes							Return/Benefits	
Program	Age	Duration	Target	Selection	Follow-Up	Sample	Home	Health	Parental	On Site	Group	IQ	School	Character	Education	Health	Crime	Earnings	Return	Benefit/Cost
Early																				
NFP	< 0	2Y	SES	Prgrm	19Y	640	☒	☒	☒	□	□	●	◐	●	○	◍	◐	.		2.9
ABC	0	5Y	SES	Refer	30Y	90	□	☒	☒	☒	☒	●	●	◐	◐	◍	◐	◐		3.8
IHDP	0	3Y	Health	Prgrm	18Y	640	☒	□	☒	☒	☒	◐	◐	◐	○	○	○	.		
FDRP	0	5Y	SES	Prgrm	15Y	110	☒	☒	☒	☒	☒	○	◐	◍	.	.	●	.		
PCDC	1	2Y	SES	Prgrm	15Y	170	☒	□	☒	☒	☒	◍	◍	●	.	.	.	.		
JSS	1–2	2Y	Health	Prgrm	22Y	160	☒	☒	☒	□	□	●	●	●	◍	●	.	●		
Perry	3	2Y	SES, IQ	Prgrm	37Y	120	☒	□	☒	☒	☒	◍	●	●	◐	○	●	◐	7–10	7.1–12.2
Head Start	3	2Y	SES	Prnt	23Y	4,170	☒	☒	☒	☒	☒	◍	◍	○	●	●	◐	●		
CPC	3–4	2Y	SES	Prnt	25Y	1,290	□	☒	☒	☒	☒	.	●	●	●	●	●	●	18	10.8
TEEP	3,5	2Y	SES	Prgrm	22Y	260	☒	☒	☒	☒	□	◍	●	●	●	.	.	.		
STAR	5–6	4Y	SES	Prgrm	22Y	11,000	□	□	□	☒	☒	.	◍	●	●	.	.	●	6.2	
Elementary																				
LA's Best	5–6	6Y	SES	Schl	12Y	19,320	□	☒	☒	☒	☒	.	◍	.	●	.	○	.		0.9
CSP	5–13	5Y	Behav	Refer	35Y	510	☒	☒	☒	☒	☒	.	.	○	.	⊗	○	.		
SSDP	6–7	6Y	Crime	Prgrm	21Y	610	□	☒	☒	☒	☒	.	●	●	●	●	◍	○		3.1
Adolescence																				
BBBS	10-16	1Y	SES	Self	1Y	960	□	□	□	□	□	.	◐	○	.	◐	○	.		1.0
IHAD	11–12	7Y	SES	Prgrm	8Y	180	□	☒	☒	☒	☒	.	.	.	●	.	.	.		
EPIS	13–15	3Y	Schl	Schl	2Y	45,070	□	□	□	☒	☒	.	●	.	.	.	.	.		0.9–3.0
xl club	14	2Y	Schl	Schl	2Y	261,420	□	□	□	☒	☒	.	○	.	.	.	.	.		
SAS	14–15	5Y	Schl, SES	Schl	6Y	430	□	□	□	□	□	.	◍	○	●	.	.	.		
STEP	14–15	2Y	Schl, SES	Self	4Y	4,800	□	☒	□	☒	☒	.	○	.	○	.	.	○		
QOP	14–15	5Y	Schl	Prgrm	10Y	1,070	□	☒	□	☒	☒	.	○	.	◍	○	⊗	○		0.42
Academies	13–16	4Y	Schl, SES	Self	12Y	1,460	□	□	□	☒	☒	.	◍	○	○	○	◍	◐		
ChalleNGe	16–18	1Y	Dropout	Self	3Y	1,200	□	☒	□	☒	☒	.	.	◍	●	◐	◍	●	6.4	2.66
Job Corps	16–24	1Y	SES	Self	9Y	15,300	□	☒	□	☒	☒	.	.	.	○	◍	◍	◍		0.22
Year-Up	18–24	1Y	SES	Self	2Y	200	□	□	□	☒	☒	.	.	.	.	.	.	●		

Notes: □ – Does not include intervention component. ☒ – Includes intervention component. ○ – No effects. ● – Positive effects. ◍ – Weakly positive effects. ◐ – Mixed effects (either different studies find different results or only particular sub-populations benefited). ⊗ – Negative effects. "." – Not measured. "Age" – The age at which participants entered the program. For programs that targeted grades, rather than ages, it was assumed that children entered kindergarten at ages 5–6 and the age range advanced one year for each subsequent grade. "Duration" – Length of the treatment. In cases where the treatment length varied for participants, the longest duration was presented. "Target" – Population that was targeted by the program. SES – socioeconomic status or disadvantage. Behav – Behavior. Schl – School Performance. Crime – local crime rates. IQ – low IQ. "Selection" – The party that acted first in joining the sample. Prgrm – Evaluation program contacted participants. Refer – Other party referred participants to program. Prnt – Parent applied to program. Self – Participant applied to program. Schl – School selected participants. "Follow-Up" – Duration of longest follow-up evaluation in years. "Sample" – Largest sample size from the studies examined (rounded to nearest 10). "Home" – Included home visits. "Health" – Included a nutritional component. "Parental" – Involved parents. "On Site" – Took place at an on site location. "Group" – Whether the intervention combined participants in groups. "IQ" – IQ score. "School" – school performance. "Character" – measured character skills. "Education" – educational attainment. "Health" – health (including drug use). "Crime" – crime. "Earnings" – earnings or related outcomes. "Return" – Annual rate of return. "Benefit/Cost" – Estimated benefits divided by costs.

Sources: **NFP** – Eckenrode, Campa, Luckey, Henderson, Cole, Kitzman, Anson, Sidora-Arcoleo, and Olds (2010); Kitzman, Olds, Cole, Hanks, Anson, Arcoleo, Luckey, Knudtson, Henderson, and Holmberg (2010); Olds (2006); Olds, Kitzman, Cole, Hanks, Arcoleo, Anson, Luckey, Knudtson, Henderson, Bondy, and Stevenson (2010); Olds, Kitzman, Cole, Robinson, Sidora, Luckey, Henderson, Hanks, Bondy, and Holmberg (2004); Olds, Kitzman, Hanks, Cole, Anson, Sidora-Arcoleo, Luckey, Henderson, Holmberg, Tutt, Stevenson, and Bondy (2007); Olds, Robinson, Pettitt, Luckey, Holmberg, Ng, Isacks, Sheff, and Henderson (2004). **ABC** – Breitmayer and Ramey (1986); Heckman, Moon, and Pinto (2012); Temple and Reynolds (2007). **IHDP** – McCormick, Brooks-Gunn, Buka, Goldman, Yu, Salganik, Scott, Bennett, Kay, Bernbaum, Bauer, Martin, Woods, Martin, and Casey (2006). **FDRP** – Lally, Mangione, and Honig (1987). **PCDC** – Besharov, Germanis, Higney, and Call (2011); Bridgeman, Blumenthal, and Andres (1981); Johnson and Walker (1991). **JSS** – Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2013); Grantham-McGregor, Powell, Walker, and Himes (1991); Walker, Chang, Powell, and Grantham-McGregor (2005); Walker, Chang, Powell, Simonoff, and Grantham-McGregor (2007). **Perry** – Heckman, Moon, Pinto, Savelyev, and Yavitz (2010a); Heckman et al. (2010b, 2013). **Head Start** – Carneiro and Ginja (2012); Currie and Thomas (1995); Deming (2009); Garces, Thomas, and Currie (2002); Ludwig and Miller (2007); Westat (2010). **CPC** – Niles, Reynolds, and Nagasawa (2006); Reynolds (1994); Reynolds and Temple (1998); Reynolds, Temple, Ou, Arteaga, and White (2011); Reynolds, Temple, Robertson, and Mann (2002); Reynolds, Temple, White, Ou, and Robertson (2011). **TEEP** – Kagitcibasi, Sunar, and Bekman (2001); Kagitcibasi, Sunar, Bekman, Baydar, and Cemalcilar (2009). **STAR** – Chetty, Friedman, Hilger, Saez, Diane, and Yagan (2011); Krueger (2003). **LA's BEST** – Goldschmidt and Huang (2007); Huang, Gribbons, Kim, Lee, and Baker (2000); Huang, Kim, Marshall, and Pérez (2005). **CSP** – McCord (1978). **SSDP** – Hawkins, Catalano, Kosterman, Abbott, and Hill (1999); Hawkins, Kosterman, Catalano, Hill, and Abbott (2005, 2008). **BBBS** – Tierney, Baldwin-Grossman, and Resch (1995). **IHAD** – Kahne and Bailey (1999). **EPIS** – Martins (2010). **XL Club** – Holmlund and Silva (2009). **SAS** – Johnson (1999). **STEP** – Walker and Vilella-Velez (1992). **QOP** – Rodríguez-Planas (2010, 2012). **Academies** – Kemple and Snipes (2000); Kemple and Willner (2008). **ChalleNGe** – Bloom, Gardenhire-Crooks, and Mandsager (2009); Millenky, Bloom, and Dillon (2010); Millenky, Bloom, Muller-Ravett, and Broadus (2011). **Job Corps** – Schochet, Burghardt, and Glazerman (2001); Schochet, Burghardt, and McConnell (2008). **Year Up** – Roder and Elliot (2011).

time, low-income, unmarried, and/or adolescent mothers without previous births. In the program, nurses visit young mothers from the first or second trimester of the mother's pregnancy until the second birthday of her first child. The program encourages mothers to reduce smoking, teaches mothers how to care for their children, and helps mothers to pursue education and find jobs. It was introduced, as a pilot, at Elmira, New York. Later, the program was introduced at Memphis, Tennessee, and Denver, Colorado.[68] At Elmira and Memphis, mothers receive around 30 to 35 visits of around 75–90 minutes each. By 2006, it operated in 26 states and served more than 20,000 families each year. It is evaluated by the method of randomized assignment (Howard and Brooks-Gunn, 2009; Olds, 2006).

NFP reduces risky behaviors of mothers and their children (see Table 3). Children exhibit persistently higher IQ scores through age 6 (Olds, Kitzman, Cole, Robinson, Sidora, Luckey, Henderson, Hanks, Bondy, and Holmberg, 2004).[69] At age 12, the children in the treatment group have lower rates of substance abuse and exhibit lower levels of internalizing behavior (e.g. anxiety, depression, and withdrawal).[70] When the children are 12, mothers are much less dependent on welfare.[71] By age 19, children are less likely to engage in crime.[72,73] Importantly, the program improves measures of criminal activity, even though there are only weak effects on achievement test scores or grades in the long run. This suggests that the program works by improving character skills.

[68]Our discussion will focus on the effects of the program up to age 12 for the Memphis sample and the effects at age 19 for the Elmira sample.

[69]Measured by the mental processing domain of the Kaufman Assessment Battery for Children (KABC).

[70]Kitzman et al. (2010). The measure of internalizing behavior is based on self-reports from the Achenbach Child Behavior Checklist (Achenbach and Rescorla, 2001). See Almlund et al. (2011) for a discussion of the relationship between internalizing behavior and the Big Five.

[71]Olds et al. (2010).

[72]Eckenrode et al. (2010).

[73]Illustrated by the strong discrepancy in Table 3 between estimates for "Ever Arrested" versus "Arrested in Last Year."

Table 3 Summary of Effects of the Nurse-Family Partnership

Outcome	**Treatment**	**Control**	**Treatment Effect**
Age 6[(1)]			
Vocabulary Skills(ES)			0.17**
Internalizing Disorders(%)	12.6	14.7	−2.1
Externalizing Disorders(%)	17.4	20.2	−2.8
Age 9[(2)]			
GPA(ES)			0.09
Antisocial Behavior(%)			−0.03
Grade Retention(%)	16.0	12.4	3.6
Age 12[(3)]			
GPA			0.08
Achievement Tests[(a)]			1.09
Grade Retention(%)	24.9	20.8	4.1
Internalizing Disorders(%)[(b)]	22.1	30.9	−8.8**
Externalizing Disorders(%)[(c)]	19.7	17.8	1.9
Used Substance Last 30 Days(%)	1.7	5.1	−3.4**
Ever Arrested(%)	3.1	3.1	0.0
Welfare Benefits Mother[(d)]	8772	9797	−1025**
Age 19[(4)]			
Ever Arrested(%)	21.3	37.4	−16.1**
Arrested in Last Year(%)	8.2	5.5	2.7
Illicit Drug Use(%)	48.7	51.9	−3.2
Has HS Diploma(%)	70.6	74.5	−3.9
Economically Productive(%)[(e)]	71.4	68.3	3.1

Notes: The estimates are coefficients from regressions that control for sample member characteristics. % refers to treatment effects in terms of changes in prevalence in outcome variable in percentage points. ES indicates Effect Size, reporting the treatment effect in standard deviations of the outcome variable. The Age 6–12 estimates come from the Memphis site. The Age 19 estimates come from the Elmira site. (*a*) Is based on group reading and math achievement test scores and is in percentile units. (*b*) Uses student self-reports on domains such as anxiety, depression, somatization, and withdrawal to assess if students pass a clinical threshold, based on the Achenbach Child Behavior Checklist (Achenbach and Rescorla, 2001). (*c*) Uses student, teacher, and parent reports on domains such as conduct problems, aggression, and total problems to assess if students pass a clinical threshold, based on the Achenbach Child Behavior Checklist. (*d*) Measures the average yearly receipt of welfare during the child's first 12 years of life, in US$. (*e*) Measures if someone is involved in education, a job, the military, or job training.
**Attains 5% significance level.
Sources:(1) The estimates at Age 6 come from Olds et al. (2004). (2) The estimates at Age 9 come from Olds et al. (2007). (3) The estimates at Age 12 come from Kitzman et al. (2010). (4) The estimates at Age 19 come from Eckenrode et al. (2010).

7.1.2 Jamaican Study

Many childhood programs have been introduced in less developed countries. The Jamaican Supplementation Study was one of the few with a long-term follow-up.[74] This randomized program consisted of a two-year nutritional and stimulation program for stunted children (low height for age), aged 9-24 months at the start of the program. Participants either received supplementation (milk formula), stimulation (encouraged mother to play with kids in an effective manner), or both (Grantham-McGregor et al., 1991).

In general, the stimulation intervention outperformed the supplementation intervention (see Table 4). Both interventions improved early cognitive development in the short-term, but only the stimulation intervention improved cognitive and character skills in the long run (Gertler et al., 2013; Grantham-McGregor et al., 1991). As with the NFP program, the stimulation program improved internalizing behavior (Walker et al., 2005). The stimulation intervention improved educational attainment and earnings. The effects on migration and earnings at age 22 are substantial (Gertler et al., 2013).[75]

[74]Most programs lack long-term follow-ups (even beyond age 10), or any follow-up at all, and few employ methods of random assignment. Short-term effects are generally positive; intervention leads to better cognitive skills and, in most cases, lower grade retention in the early years of primary school. For an overview, see Engle, Fernald, Alderman, Behrman, O'Gara, Yousafzai, Cabral de Mello, Hidrobo, Ulkuer, Ertem, and Iltus (2011); Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, and Strupp (2007); Grantham-McGregor et al. (1991); Jolly (2007); Lake (2011); Myers (1992); Walker, Wachs, Grantham-McGregor, Black, Nelson, Huffman, Baker-Henningham, Chang, Hamadani, Lozoff, Meeks Gardner, Powell, Rahman, and Richter (2011); Walker et al. (2005); Walker, Chang, Vera-Hernández, and Grantham-McGregor (2011); Walker, Wachs, Gardner, Lozoff, Wasserman, Pollitt, Carter, and The International Child Development Steering Group (2007); Young (1996) Engle, Black, Behrman, Cabral de Mello, Gertler, Kapiriri, Martorell, Eming Young, and The International Child Development Steering Group (2007).

[75]The supplementation program might have been ineffective in the long run because the supplement was relatively weak, contained few micronutrients, and had no fortified formula. Some of the interventions referred to in Myers (1992) and Young (1996) also reported stronger effects of stimulation over supplementation.

Table 4 Summary of Treatment Effects from the Jamaican Study

Outcome	**Supplementation**	**Stimulation**
Age 2$^{(1)}$		
Mental Development(ES) $^{(a)}$	0.76***	0.91***
Age 17–18$^{(2)}$		
Full Scale IQ(ES)	−0.03	0.58**
Performance IQ(ES)	0.04	0.51*
Verbal IQ(ES)	−0.05	0.50**
Reading Score(ES)$^{(b)}$	−0.05	0.66***
Internalizing Behavior(ES)		0.39**
Externalizing Behavior(ES)		0.22
Age 20–22$^{(3)}$		
Years of Education		0.66
Currently in Education(ES)		0.14*
Log Earnings of Last Job		0.33**
Employment(%)		8.0

Notes: The estimates are coefficients from regressions that control for sample member characteristics. Grantham-McGregor et al. (1991) and Walker et al. (2005) test and do not reject the hypothesis that "stimulation only" and "supplementation only" treatment effects add to produce the "both" treatment effect. We report the effects of separate treatment streams: "stimulation only" and "supplementation only." Gertler et al. (2013) do not use the supplementation group at all and pool the stimulation and "both" groups into a single treatment group. % refers to treatment effects in terms of changes in prevalence in outcome variable in percentage points. ES indicates Effect Size, reporting the treatment effect in standard deviations of the outcome variable. (a) Assessed with Griffiths Mental Development Scales (Griffiths, 1954). The values have been divided by the standard deviation of the control group. (b) The reading score is based on the sentence completion portion of the Group reading test 2, revised. The effect is divided by the standard deviation of the control group.

* 10% significance;** 5% significance; *** 1% significance.

Sources: (1) The estimates at Age 2 come from Grantham-McGregor et al. (1991). (2) The estimates at Age 17–18 come from Walker et al. (2005) and Gertler et al. (2013). (3) The estimates at Age 20–22 come from Gertler et al. (2013).

7.1.3 Perry Preschool Program

Evaluations of the Perry Preschool program provide some of the most compelling evidence that character skills can be boosted in ways that produce adult success. The Perry Preschool program enriched the lives of 3- and 4-year-old low-income black children with initial IQs below 85 at age three.[76] They attended 2.5 hours of center-based preschool five days a week for two years. In addition, home visits promoted parent-child relationships. The program ended after two years of enrollment, and both treatment and control groups entered the same school.

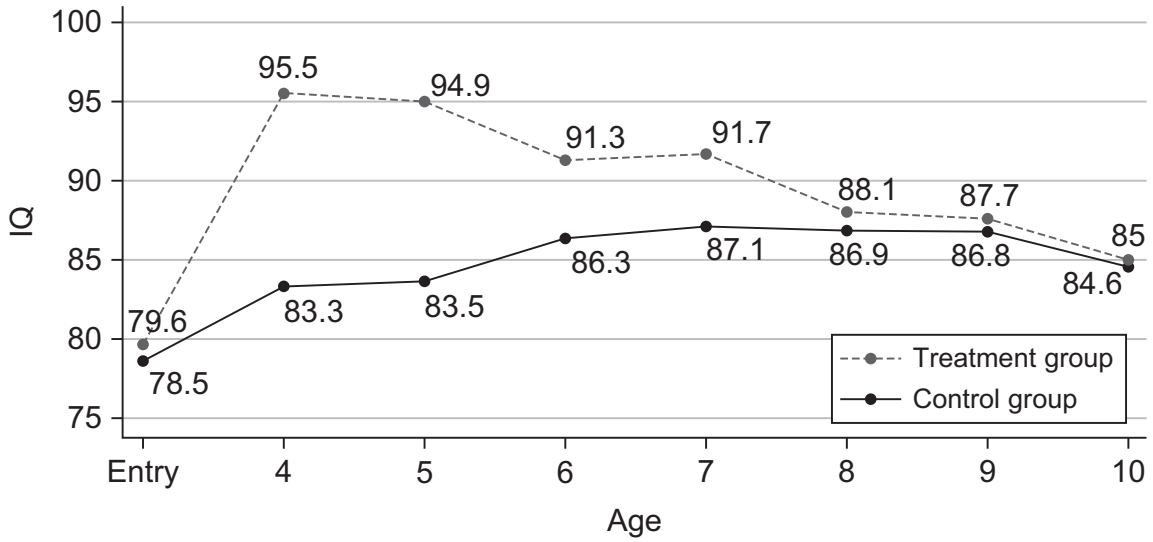
Participants were taught social skills in a daily "plan-do-review" sequence in which children first planned a task, executed it, and then reviewed it with teachers and fellow students. They learned to work with others when problems arose.[77] The program was evaluated by the method of random assignment.

The program did not improve IQ scores in a lasting way. Figure 10 shows that, by age 10, treatment and control groups had the same average IQ scores. Many critics of early childhood programs seize on this finding and related evidence to dismiss the value of early intervention studies.

[76] We draw on the analysis of Heckman et al. (2010a,b, 2013).

[77] Sylva (1997) describes the Perry program as a program that fosters character skills. It has features in common with the Tools of the Mind intervention discussed below. See Bodrova and Leong (2001) for the Tools of the Mind curriculum. See Heckman and Pinto (2013) for a discussion of the Perry curriculum and the Tools of the Mind curriculum.

Figure 10 Perry Preschool Program: IQ, by Age and Treatment Group



Source: Cunha et al. (2006) and Heckman and Masterov (2007), based on data provided by the High Scope Foundation.
Notes: IQ measured on the Stanford-Binet Intelligence Scale (Terman and Merrill, 1960). The test was administered at program entry and at each of the ages indicated.

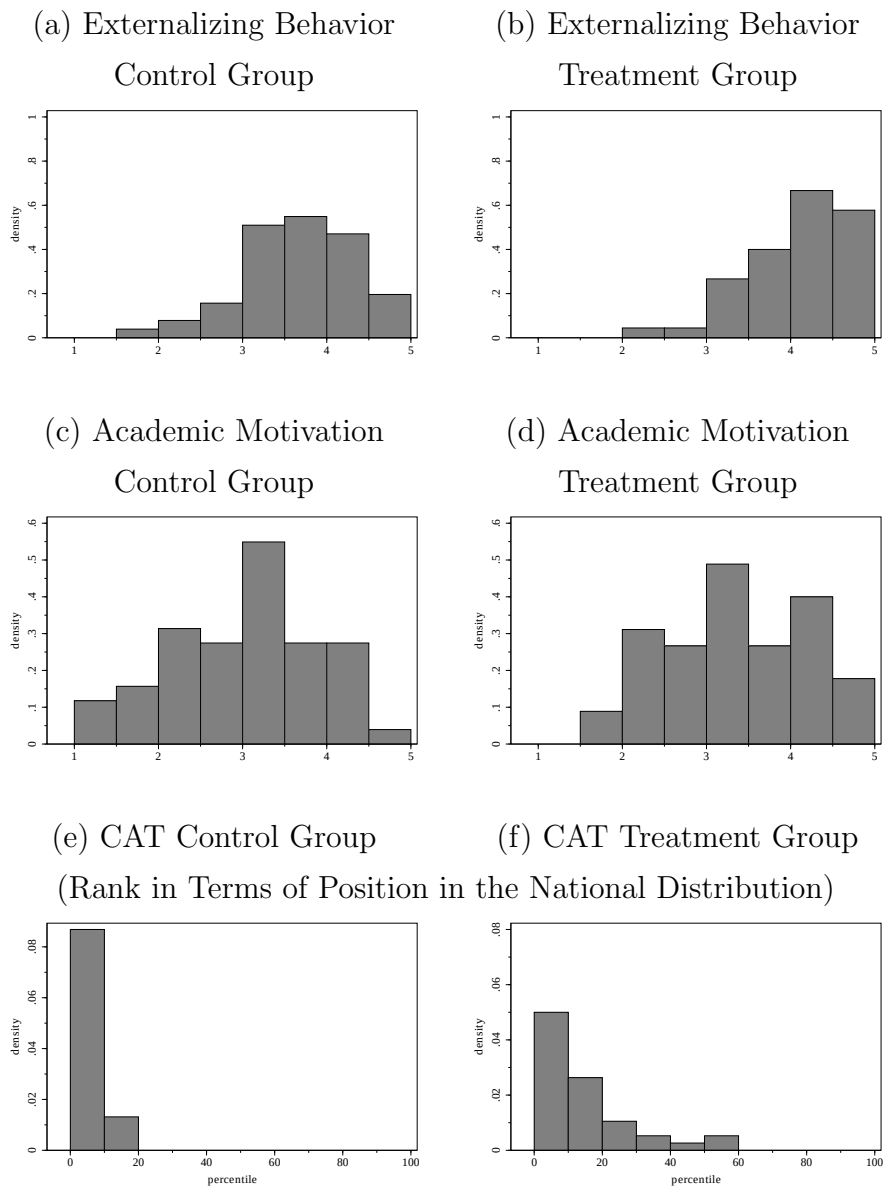
Nevertheless, the program improved outcomes for both boys and girls, resulting in a statistically significant rate of return of around 7%–10% per annum for both boys and girls (see Heckman et al., 2010b). These returns are in the range of the post-World War II, pre-2008 meltdown stock market returns to equity in the U.S. labor market, estimated to be 6.9% per annum.[78]

Heckman, Pinto, and Savelyev (2013) show that the Perry treatment effects arise primarily from lasting changes in character skills, not from changes in IQ. Figure 11 shows histograms of measures of character skills for the treatment and control groups. The treatment groups of both genders improved their teacher-reported externalizing behavior, a skill related to Agreeableness and Conscientiousness. For girls, the program also improved Openness to Experience (proxied by academic motivation). Figure 11 also shows that the program improved scores on the California Achievement Test (CAT), even though it did not have a

[78]See DeLong and Magin (2009).

lasting effect on IQ. This evidence is consistent with the evidence summarized in Figures 2 and 3 showing that achievement test scores depend strongly on character skills (see Borghans et al., 2008; Borghans, Golsteyn, Heckman, and Meijers, 2009). Achievement tests measure general knowledge. The acquisition of general knowledge (crystallized intelligence) depends on persistence, curiosity, and focus.

Figure 11 Histograms of Indices of Noncognitive Skills and CAT Scores



Source: Heckman et al. (2013).

Notes: Indices for Externalizing Behavior and Academic Motivation are based on items of the Pupil Behavior Inventory (PBI), teacher ratings of student behavior. The units are expressed in terms of standard deviations. The scale on "Externalizing Behavior" is normalized so that a higher score corresponds to better behavior. The PBI includes whether the student disrupts classroom. "CAT" is the California Achievement Test score expressed in percentiles of the general population distribution of the scores. The one-sided p-values for difference in means are 0.001, 0.043, and 0.000 for Externalizing Behavior, Academic Motivation, and CAT scores, respectively. Histograms are based on the pooled sample of males and females.

7.1.4 Abecedarian Program

The Abecedarian (ABC) program was another program aimed at disadvantaged black children. It started earlier, lasted much longer, and was much more intensive than the Perry program, combining a preschool intervention that started when children were as young as 6 weeks old with school-age treatment through grade three.[79] The preschool component was full-day child care five days per week, 50 weeks a year. The curriculum focused on a series of "educational games" that emphasized language, emotional development, and the development of cognitive skills. It also had a medical and nutritional component.[80] For grades one through three, the curriculum was tailored to the individual strengths of the students. During this phase, teachers and parents interacted on a bi-weekly basis.

In contrast to Perry, ABC led to lasting improvements in IQ.[81] For girls, the program improved IQ through age 21 (see Figure 12). The effect for boys was positive but was less precisely estimated. Girls and boys also scored better on achievement tests. ABC likely improved IQ because it started at an earlier age than Perry. Very early childhood appears to be a critical period for shaping IQ.

As with Perry, the benefits of the ABC program differ across genders. For girls, the program improved educational attainment, reduced participation in criminal activity, decreased substance abuse, and improved internalizing and externalizing behavior. Like the Perry program, ABC improved employment and health for males and produced substantial improvements in character skills.[82]

[79] It is easy to exaggerate the difference in intensity between the Perry program and the ABC program. While the ABC program starts earlier and spends more time each day with the child, the amount of time per day spent on essentially the same type of learning activities is very similar at ages 3 and 4. See Griffin, Heckman, and Moon (2013).

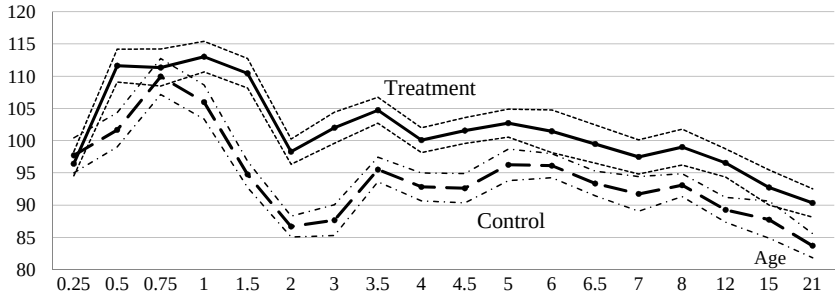
[80] See Campbell, Pungello, Miller-Johnson, Burchinal, and Ramey (2001); Campbell, Ramey, Pungello, Sparling, and Miller-Johnson (2002).

[81] We report on the treatment effects of the preschool component from Heckman et al. (2012).

[82] Campbell, Conti, Heckman, Moon, and Pinto (2013) report substantial beneficial effects of the program on adult health for both genders.

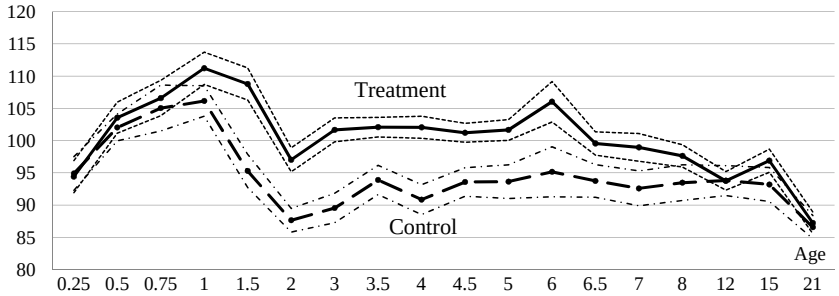
Figure 12 Cognitive Tests by Gender and Daycare Treatment Status

(a) Females



	0.25	0.5	0.75	1	1.5	2	3	3.5	4	4.5	5	6	6.5	7	8	12	15	21
Treatment	96.4	111.6	111.3	113.0	110.4	98.3	102.0	104.7	100.1	101.6	102.7	101.5	99.5	97.5	99.0	96.5	92.7	90.3
Control	97.7	101.7	109.9	106.0	94.8	86.7	87.7	95.5	92.8	92.6	96.3	96.1	93.4	91.7	93.1	89.3	87.7	83.7

(b) Males



	0.25	0.5	0.75	1	1.5	2	3	3.5	4	4.5	5	6	6.5	7	8	12	15	21
Treatment	94.4	103.6	106.6	111.2	108.8	97.0	101.7	102.1	102.1	101.2	101.7	106.0	99.5	99.0	97.6	93.8	96.9	87.2
Control	94.9	102.0	105.0	106.1	95.3	87.6	89.5	93.9	90.9	93.6	93.6	95.2	93.7	92.6	93.5	93.8	93.2	86.6

Source: Heckman et al. (2013).
Notes: All the cognitive tests have been standardized to population mean of 100 and standard deviation of 15. The measures used in this figure are Bayley Mental Development Index at age 0.25, 0.5, 0.75, 1, and 1.5; Stanford Binet Intelligence Scale at age 2, 3, 4, 5, and 6; McCarthy Index at age 3.5, 4.5, and 7; WPPSI IQ Score at age 5; WISC-R Intelligence Scale at age 6.5, 8, 12, and 15; WAIS-R Adult Intelligence Scale at age 21.

7.2 Large-Scale Infant and Young Child Programs

7.2.1 Head Start

The Head Start program receives considerable attention both in the evaluation literature and in public discussion. Head Start children are eligible for enrollment from ages three to five, although they generally enter at age four and receive one year of treatment. In addition to a center-based (preschool) intervention, Head Start includes medical services and parental assistance (Ludwig and Miller, 2007). Its implementation differs greatly across sites, making it difficult to evaluate its overall effectiveness (Deming, 2009). Parental participation was not mandatory, and there is some evidence that many do not participate.[83]

There have been no long-term evaluations of Head Start based on randomized assignment. Instead, most evaluations of Head Start compare siblings with different enrollment statuses[84] or use regression discontinuity designs.[85] Empirical evidence on Head Start is mixed. Many studies find that Head Start improves IQ scores and achievement test scores, but that the improvements fade by age 10 (Deming, 2009; Ludwig and Miller, 2007). For some subgroups, there is evidence of persistent gains in test scores and reductions in retention rates.[86] Garces et al. (2002) find that Head Start improves outcomes for educational attainment for whites and reduces criminality for blacks, whereas Deming (2009) finds that Head Start improves educational attainment for blacks.

A short term experimental study by Westat (2010) based on randomized assignment offers a more pessimistic view. The study finds that most positive impacts on achievement tests[87]

[83] Schumacher (2003)

[84] For example, Deming (2009); Garces et al. (2002).

[85] For example, Carneiro and Ginja (2011); Ludwig and Miller (2007).

[86] Deming (2009) reports persistent gains for males and those with higher ability mothers as measured by summary test score that includes the Peabody Picture Vocabulary Test (PPVT), Peabody Individual Achievement Math (PIATMT), and Reading Recognition (PIATRR). Retention rates were lower for boys, blacks, and students with low maternal AFQT.

[87] Including the Peabody Picture Vocabulary Test and several batteries of the Woodcock-Johnson III test.

and socio-emotional development[88] fade by first grade. This evaluation likely underestimates the true program effects. Many members of the control group joined the program at sites other than where they initially applied or were enrolled in other childhood programs that were more intensive than Head Start.

7.2.2 Chicago CPC

The Chicago Child–Parent Center (CPC) is an apparently successful large-scale preschool program. It is directed toward disadvantaged, predominantly black, inner-city children in Chicago. The program provides a half- or full-day preschool program for 3- and 4-year-olds. The program was designed to develop basic reading, writing, and math skills. In contrast to other large-scale programs, parents are required to visit the centers and receive advice on good parenting behavior (Reynolds, 2000).

CPC has not been evaluated by the random assignment method. All evaluations are based on samples of children attending CPC kindergarten matched to "comparable" children attending non-CPC kindergarten. Based on these methods, the program appears to improve education and reduce criminal behavior (see Table 5). Although it is difficult to tell what causes the improvement in adult outcomes, some evidence suggests that boosts in character play a role. At age 13, the participants have higher levels of social and emotional competence. The program also improves achievement test scores (Iowa Tests of Basic Skills),[89] but improvements in character can lead to improvements in achievement tests without affecting IQ, as was the case with the Perry program.

CPC might have been more successful than other similar programs, like Head Start, because it encouraged parental involvement. Parents are induced to visit the centers, and the staff teaches them various forms of stimulative play that they can do at home and assists

[88] Including parent- and teacher-reported measures of aggression, hyperactivity, withdrawn behavior, and total problem behavior. They use several test batteries, including the Achenbach Classroom Behavior Checklist (CBCL; Achenbach and Rescorla, 2001) and the Adjustment Scales for Preschool Intervention (ASPI; Lutz, Fantuzzo, and McDermott, 2000).

[89] Hieronymus, Lindquist, and Hoover (1980).

parents interested in pursuing further education and seeking jobs (Reynolds, 2000).

Table 5 Summary of Effects from the Chicago Child-Parent Center

Outcome	**Treatment**	**Control**	**Treatment Effect**
Age 13$^{(1)}$			
Reading Achievement(ES)			0.36***
Math Achievement(ES)			0.32***
Social and Emotional Competence(ES)			0.19*
Problem Behavior(ES)			−0.17
Grade Retention(ES)			−0.3***
Age 18–20$^{(2)}$			
Has HS Diploma(%)	49.7	38.5	11.2***
Number of Arrests	0.45	0.78	−0.33**
Number of Violent Arrests	0.22	0.35	−0.13**
Age 22–26$^{(3)}$			
Has HS Diploma(%)	79.7	72.9	6.8***
Number of Adult Felony Arrests	0.32	0.44	−0.12**
Substance Abuse(%)	14.3	18.8	−4.5**
Full-time Employment (%)	36.8	33.5	3.3
Age 28$^{(4)}$			
On Time HS graduation(%)	44.3	36.6	7.7**
Any Adult Arrest(%)	47.9	54.3	−6.4**
Substance Abuse(%)	13.7	18.9	−5.2**
Annual Income($)	11,582	10,796	786**

Notes: The estimates are coefficients from regressions that control for sample member characteristics. Estimates apply to the pre-school part of CPC, excluding extended treatment. ES indicates Effect Size, reporting the treatment effect in standard deviations of the outcome variable.
* 10% significance;** 5% significance; *** 1% significance.
Sources: (1) The estimates at Age 13 come from Reynolds (1995) and Niles et al. (2006). (2) The estimates at Age 18–20 come from Reynolds, Temple, Robertson, and Mann (2001). (3) The estimates at Age 22–26 come from Reynolds et al. (2011). (4) The estimates at Age 28 come from Reynolds et al. (2011).

8 Education and Interventions in Kindergarten and Elementary School

150 years ago, moral education, rooted in Protestant Christianity, was taught in American public schools.[90] Moral and character education disappeared from the curriculum of public schools. However, moral education is featured in Catholic education. Some private schools teach a secular version of moral education. For example, the Knowledge Is Power Program (KIPP) Charter Schools enforce strict codes of discipline and give character training to their participants (The KIPP Foundation, 2011). In contemporary society, discussions of moral and character education provoke controversy even among its advocates. Scholars disagree about the origins of character and morality and how they can best be fostered.[91]

In this section, we survey the evidence on secular versions of moral and character education. This does not reflect any antireligious bias on our part. Instead, it reflects the absence of convincing evaluations of the effectiveness of character education/moral development programs in religious schools.

8.1 Targeted Character Interventions

Few well-evaluated interventions have targeted children in elementary school. Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) present a meta-analysis of 213 school-based social and emotional learning programs. While their study suggests that these programs have been successful, many of the evaluations they review suffer from substantial methodological problems. For example, only 15% of the studies have follow-ups that last beyond 6 months. For the studies with longer follow-up, the mean impact is positive and statistically signif-

[90] See the discussions in Heckman and Kautz (2014a) and Kaestle (1984).

[91] The conceptualizations of morality and character differ in the literature. Character is usually viewed as a skill acquired through habituation and practice. In its perfected state it becomes a gut reaction. Morality is viewed as a consequence of conscious choices. Many dismiss discussions of morality and character out of hand because of their religious connotations or because they suggest a prudish vision of society that does not tolerate cultural pluralism. Lapsley and Yeager (2012) survey the controversies in the field, the curricula available and evidence on their effectiveness.

icant. However, these follow-ups are still short compared to those for the Perry Preschool program and ABC.[92] Although these programs appear promising, evidence on their long-run effectiveness is lacking.

8.1.1 The Seattle Social Development Project (SSDP)

Of all of the programs reported by Durlak et al. (2011), the Seattle Social Development Project (SSDP) has been evaluated the most thoroughly. The program targeted public elementary schools in high-crime areas of Seattle. The full program lasted from first grade through sixth grade.

The program emphasized attachment between children and their parents and teachers. Throughout elementary school, the participants' teachers received five days of training per year that included proactive classroom management, interactive teaching, and cooperative learning. The first-grade teachers received additional lessons based on the Cognitive Problem-Solving curriculum, which teaches children to resolve conflicts with peers (Shure and Spivack, 1988). In first and second grade, parents were offered a seven-session course on behavioral management training.[93] In second and third grade, parents were offered a four-session course to support their children's academic achievement.[94] Finally, in fifth and sixth grade, parents were offered a five-session course designed to reduce drug use.[95] Parents of 43% of the children participated in parenting classes.[96]

The program improved a variety of long-run outcomes. Table 6 shows that SSDP improved grades and behavior during adolescence. At age 21, participants were 10 percentage points more likely to have graduated from high school (or have earned an equivalency de-

[92]The largest federal study to date on character education programs failed to find evidence for improvements in behavior or academic performance (see Social and Character Development Research Consortium, 2010).

[93]The course followed the "Catch 'Em Being Good" curriculum (Hawkins, Von Cleve, and Catalano Jr, 1991).

[94]The course consisted of four sessions that followed the "How to Help Your Child Succeed in School" curriculum (Abbott, O'Donnell, Hawkins, Hill, Kosterman, and Catalano, 1998).

[95]The course followed the "Preparing for the Drug (Free) Years" curriculum (Haggerty, Kosterman, Catalano, and Hawkins, 1999).

[96]Hawkins et al. (1999).

gree). By age 24, this effect faded, but this likely occurred because members of the control group earn GED certificates.[97] GED certificates do not confer the same benefits as a traditional high school diploma.[98] By ages 24 and 27, the participants were 12 percentage points more likely to earn an associate's degree. The treatment group had higher earnings at ages 24 and 27, although treatment effects on income are not precisely determined. The program improved self-efficacy at age 21 and 24 but did not have a statistically significant effect on achievement test scores. The program improved mental health in the long run. Like many other programs, if one judged it solely on the basis of achievement tests, SSDP would be unsuccessful. Considering meaningful life outcomes and character outcomes presents a far more positive picture.

[97]Few people earn a traditional high school degree between age 21 and 24.

[98]See Heckman, Humphries, and Kautz (2014a) and the literature they review.

Table 6 Summary of Treatment Effects from the Seattle Social Development Project

	Age			
Outcome	Age 18[h]	Age 21[i]	Age 24[j]	Age 27[j]
GPA	0.24*			
CAT(ES)[a]	0.05			
Grade Repetition(%)	−8.7**			
Dropout(%)	−7.3			
School Misbehavior[b]	−1.41**			
Violent Crime(%)	−11.4**			
Ever Arrested(%)	−6.0			
Arrested past year(%)		−2.0	1.0	1.0
Smoking(%)	−0.7			
Pregnancy(%)	−9.3*		−9.0	−8.0
Anxiety(%)[c]		−2.0		
Depression(%)[c]		−8.0*		
High School Graduate/GED(%)		10.0***	6.0	6.0
More than 2 Years of College(%)		8.0***		
Self-efficacy[d]		0.17***	0.13*	−0.01
Associate's Degree(%)			12.0*	12.0*
Bachelor's Degree(%)			7.0	6.0
Substance Abuse Index[e]			3.0	−3.0
Mental Health Disorder Index[f]			−9.0*	−11.0**
Income(in thousands)[g]			3.51	3.12

Notes: The estimates are coefficients from regressions that control for sample member characteristics. % refers to treatment effects in terms of changes in prevalence in outcome variable in percentage points. ES indicates Effect Size, reporting the treatment effect in standard deviations of the outcome variable. (*a*) CAT stands for California Achievement Test score and combines reading, language, and mathematics subtests. It has been standardized based on the sample of ninth-grade Seattle students. (*b*) Measures frequency of occurrence of skipping, cheating, and being sent from class. (*c*) Anxiety, social phobia, and depression were measured using the *DSM-IV* (American Psychiatric Association, 1994). The mental health disorder index groups this for anxiety, social phobia, posttraumatic stress disorders, and major depressive episodes. (*d*) Measured as mean score on six items concerning perceived future opportunities, on a scale of 1–4. (*e*) The Substance Abuse Index measures dependence on substances (tobacco, alcohol, and illicit drugs) using *DSM-IV* criteria. (*f*) The Mental Health Disorder Index summarizes problems of anxiety, social phobia, posttraumatic stress and depression, using *DSM-IV* criteria. (*g*) Refers to income from all sources, before taxes. Includes zero-earners; income is top-coded at $200,000. (*h*) Hawkins et al. (1999). (*i*) Hawkins et al. (2005). (*j*) Hawkins et al. (2008).
* 10% significance; ** 5% significance; *** 1% significance.

8.1.2 Cambridge-Somerville Program

The Cambridge-Somerville Program is an example of a primary school program that harmed participants. This five-year intervention targeted five to thirteen year-old boys with behavioral problems. The program was evaluated by the method of random assignment. The treatment among participants differed, but included medical assistance, tutoring, access to summer camps, and meetings with parents (McCord and McCord, 1959). A 30-year follow-up showed that participants exhibited worse drinking habits, more serious mental diseases, more heart problems, higher blood pressure, (modestly) higher crime rates, and died younger (McCord, 1978). Despite the negative effects, participants reported that they thought the program helped them. McCord (1978) speculates that the program failed because it did not create a sense of autonomy among participants. It gave intensive assistance to participants. When the assistance was removed, they fell back into unfavorable behaviors. These were aggravated by feelings of rejection and resentment for lack of support.

8.1.3 Project STAR

Project STAR provides evidence that higher quality kindergarten classes improve later-life outcomes by improving character. From kindergarten through third grade, children and teachers were randomly assigned to classrooms of differing class sizes. In an analysis that builds on the work of Heckman et al. (2013), Chetty et al. (2011) examine the Project STAR program and find that students placed in higher quality kindergarten classes tended to have higher test scores at the end of kindergarten. They measure quality by peer average performance on the Stanford Achievement Test. The effect on test scores faded over time. By eighth grade, students in better kindergarten classes performed no differently on tests.

As with the Perry program, benefits emerge later in life. Children placed in better kindergarten classrooms had significantly higher earnings in early adulthood. Furthermore, kindergarten classroom quality also predicts better fourth- and eighth-grade behavior as measured by teacher-assessed effort, initiative, interest in the class, and disruptive behavior.

In turn, measured behavior predicts earnings in adulthood, suggesting that improvement in character skills is the main channel through which better kindergarten classrooms improve earnings. Several studies suggest that smaller class sizes per se do not improve long-run outcomes.[99]

The analyses of the Perry Preschool program and Project STAR highlight the importance of long-term follow-ups and measuring outcomes that depend on character skills. If these studies had only measured short-run improvements in cognitive ability, the programs would appear to have had little effect.[100]

9 Education and Interventions Targeted Toward Adolescents and Young Adults

Compared to early-childhood programs, the evidence on adolescent programs is less abundant. There are, however, some well-documented and promising interventions. We divide adolescent programs into four main categories: (1) mentoring programs for at-risk students; (2) residential-based education programs for high school dropouts; (3) in-school, professional training; and (4) incentives for student performance. Several mentorship programs and residential-based programs appear promising but have relatively short follow-ups. Somewhat similar programs with long follow-ups have had no long-term effects. These programs appear to achieve their short-term effects by closely monitoring participants and removing them from their usual environments.[101] Programs that combine work and education are more promising and have been shown to have lasting effects. A few programs that combine long-term mentoring with financial support to attend college have improved educational attainment but do not measure other outcomes.

[99] Chetty et al. (2011); Hanushek (2003).

[100] However, the estimates reported in Chetty, Friedman, Olsen, and Pistaferri (2011) appear to be implausibly large for the effect of one teacher in one class on the lifetime earnings of students.

[101] These are called "incapacitation" effects in criminology.

Few programs that target adolescents have long-term follow-ups. As Table 2 shows, only two evaluations of adolescent programs have follow-ups of at least 10 years, compared to 14 early-childhood and elementary school evaluations. Most adolescent intervention programs tend to measure fewer outcomes and focus mainly on schooling and employment.

Our framework on the life-cycle development of skills provides an interpretation of the relative failure of adolescent programs. Early childhood is a sensitive period for the development of both cognitive and character skills—prevention is more effective than remediation. Early investments are productive because early skills promote the development of later skills and, through dynamic complementarity, make later investments more productive. The finding that short-term benefits fade away quickly in most adolescent interventions might indicate the importance of incentives and situations in affecting behavior. As previously noted, several adolescent programs temporarily supervise and control the environments of adolescents, thereby changing their incentives which beneficially, but temporarily, affects their behavior.

9.1 Adolescent Mentorship Programs

9.1.1 Quantum Opportunity Program

The Quantum Opportunity Program (QOP) is one of the few adolescent mentorship programs with a long-term follow-up. Rodríguez-Planas (2012) estimates the effect of the QOP on labor market outcomes, educational attainment, and risky behaviors. The program was an intensive after-school program that provided mentoring, educational services, and financial incentives during the four years of high school (plus one additional year in case students fell behind). The program targeted low-performing high school students. QOP was evaluated by the method of random assignment.

QOP was not successful in improving risky behaviors, and, if anything, caused harm for males. Table 7 summarizes the effects of the program. Ten years after the start of the program, males in the treatment group were *more* likely to have been recently arrested. Males were more likely to have completed two years of college or training, but the increased

educational attainment did not translate into higher earnings in the follow-up period. Female participants had better educational outcomes during the first follow-up near the end of the program, but the control group caught up by the end of the three-year follow-up.

Because the program had many components, it is difficult to pinpoint why the program failed. Rodríguez-Planas (2010) suggests four reasons. First, increasing awareness about the consequences of risky behavior could have caused teens to become fatalistic; for example, some might feel that drug addiction was inevitable. Second, the mentors might have served as advocates when students were in trouble at school or with the law, lowering the cost of engaging in problem behaviors. Third, students might have used their stipends to purchase drugs or alcohol. Fourth, the program organized group activities among the participants, which could have induced negative peer-effects.

Table 7 Summary of Treatment Effects from the Quantum Opportunity Program

	Males			Females		
Outcome Since Assignment	5 Y	7 Y	10 Y	5 Y	7 Y	10 Y
Education and Employment						
High School Diploma (%)	3.5	3.2	2.9	7.4*	3.7	0.7
Obtained GED(%)	−2.2	1.2	1.7	1.2	0.7	−0.7
Math Test Score (percentile)$^{(a)}$	0.7	–	–	0.2	–	–
Reading Test Score (percentile)$^{(a)}$	0.7	–	–	0.2	–	–
Ever in College	–	7.3	7.4	–	5.1	2.0
Completed 2 Years of College or Training	–	–	7.7*	–	–	6.2
Earned BA	–	–	0.3	–	–	2.4
Has a Job	–	−6.5	−2.9	–	0.2	5.5
Hourly Wage$^{(b)}$	–	−1.2*	−1.3	–	0.8	0.2
Annual Earnings$^{(b)}$	–	–	230	–	–	1,397
Risky Behaviors and Welfare						
Binge Drinking in Past 30 Days	5.0	−10.1**	−2.4	−0.9	−0.5	3.5
Used Illegal Drugs in Past 30 Day	5.9	−6.3	−2.2	6.7	−2.0	1.9
Arrested or Charged$^{(c)}$	−8.3*	1.8	7.8**	2.7	−2.2	0.2
Receiving Welfare	–	5.8*	4.5	–	−1.0	1.5

Notes: The estimates show the difference in means between the treatment and control groups, regardless of whether they participated in QOP. Weights were used to account for nonresponse and survey design. The years at in the columns indicate time since the start of the program. (a) The tests were developed from National Education Longitudinal Study (NELS) tests and scored by the Educational Testing Service (ETS). (b) Nonworkers were assigned values of "0." (c) For the 5-year follow-up, the value indicates percentages of ever being arrested or charged. For the 7-year follow-up, the value indicates percentages over the past 3 months. For the 10-year follow-up, the value indicates percentages over the past 2 years.
* 10% significance;** 5% significance; *** 1% significance.
Source: Rodríguez-Planas (2012).

9.1.2 Big Brothers Big Sisters

The Big Brothers Big Sisters (BBBS) mentoring program effectively improves educational outcomes (Tierney et al., 1995). The program is aimed at children (ages 10-16) living in unstable family environments, which generally are single-parent households. Volunteer mentors have regular and lengthy meetings with the enrollees for an average treatment length of one year. Mentors have flexibility in their guidance, but the program emphasizes five primary goals: developing a successful relationship; providing social, cultural, and recreational enrichment; improving peer relationships; improving self-concept; and improving motivation.

Tierney et al. (1995) evaluate the program by comparing treated children to a control group that was put on a waiting list for the program. As Table 8 shows, BBBS has positive impacts on academic outcomes but only for girls. The effects on direct measures of character skills related to Big Five Neuroticism (Self-Worth, Social Acceptance, and Self-Confidence) are positive but not statistically significant. The treatment group children are less likely to hit other children or lie to their parents. The follow-up lasted only 18 months after random assignment. Given the results from QOP, this follow-up is too short to determine whether the effects are persistent. However, BBBS differed in important ways from QOP. BBBS features one-on-one mentorship, which allows mentees and mentors to form stronger attachments and might help avoid the negative peer effects of grouping at-risk youth together.

Table 8 Summary of Effects from Big Brother Big Sister Program 18 Months after Assignment

	Treatment Effects	
Outcome	Males	Females
Start Drug Use (Odds Ratio)	0.58^{**}	0.77
Start Alcohol Use (Odds Ratio)	0.83	0.68
Number of Times Hit Someone	-0.67	-1.17^{*}
Number of Times Stole Something	-0.07	-0.02
Number of Times Damaged Property	-0.04	-0.03
Perceived Ability to Complete Schoolwork	0.39	1.25^{***}
Grade Point Average	0.03	0.17^{**}
Number of Times Skipped Class	-0.18	-1.07^{***}
Number of Times Lied to Parent	-0.83	-2.24
Global Self-Worth$^{(a)}$	0.24	0.37
Social Acceptance$^{(a)}$	0.54	0.09
Self-Confidence$^{(b)}$	0.01	0.46

Notes: The estimates come from models that control for baseline characteristics. The effects for "Start Drug Use" and "Start Alcohol Use" were estimated using a logit model. All other effects were estimated using ordinary least squares. (a) From "Self-Perception Profile for Children" (Harter, 1985). (b) From "Self-Image Questionnaire for Young Adolescents" (Petersen, Schulenberg, Abramowitz, Offer, and Jarcho, 1984).
* 10% significance; ** 5% significance; *** 1% significance.
Source: Tierney et al. (1995).

9.1.3 EPIS Program

Martins (2010) analyzes data from EPIS, a program developed to improve student achievement of 13- to 15-year-olds in Portugal. Unlike many other interventions, EPIS emphasizes character skills, aiming to increase motivation, self-esteem, and study skills. The program consists of one-on-one meetings with a trained staff member or meetings in small groups. The intervention is tailored to each participant's individual skill deficit. EPIS decreases grade retention by 10 percentage points, in a cost-effective way. Unfortunately, other outcomes are not reported.

9.2 Residential-Based Programs

9.2.1 Job Corps

The Job Corps is the largest residential training program in the United States for at-risk youth. Youth spend one year in training and remedial education, and receive counseling and training in social skills. About 90% of participants live at a residential facility throughout the program (Schochet et al., 2001). Early evaluations found that the program improved many outcomes including wages, welfare dependence, college attendance, health, and crime.[102] However, more recent evaluations report that gains in earnings fade away rather quickly and that the reduced crime effects came primarily during the program and only for the residential participants (Schochet et al., 2001, 2008). Figure 13 shows that net earnings effects are basically zero.[103]

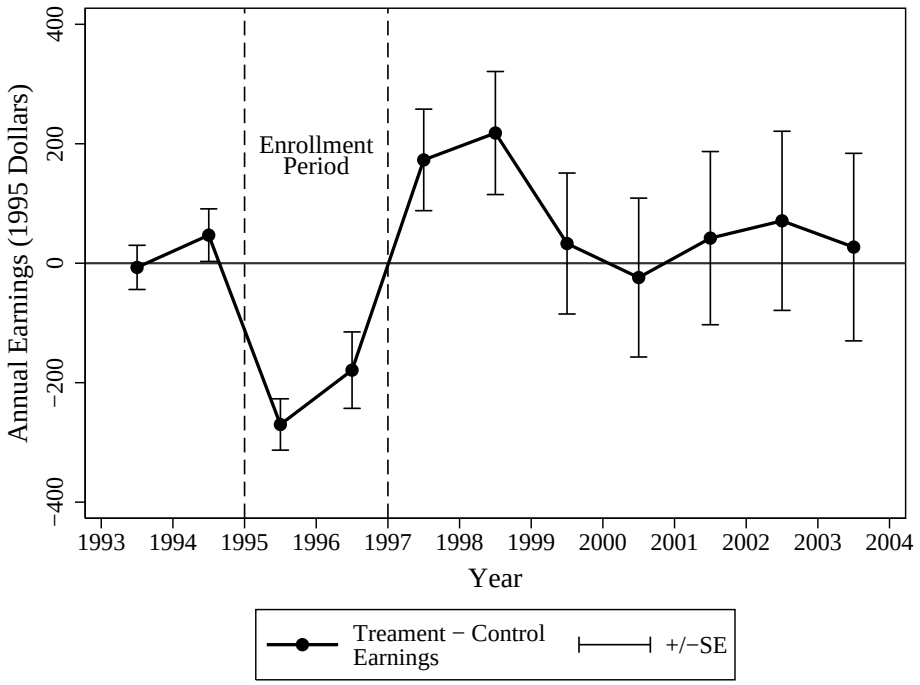
This program provides useful insights into the ineffectiveness of achievement tests to predict life outcomes. The Job Corps promoted GED certification, which is based on passing an achievement test. Early, optimistic evaluations of the program (Schochet et al. (2001)) treated GED recipients as equivalent to high school graduates and projected substantial

[102]See, for example, Mallar, Kerachsky, Thornton, and Long (1982).

[103]For males ages 20–24 there appear to be modest Job Corps effects. However, these findings do not survive adjustments for multiple hypothesis testing (Schochet et al. (2008)).

earnings gains. Later evaluations based on realized earnings showed the ineffectiveness of the Job Corps and implicitly the GED.

Figure 13 The Treatment Effect of Job Corps on Annual Earnings



Source: Estimates from Schochet et al. (2008).
Notes: The values in the graph are the difference between the annual earnings in the treatment group and the control group. The dashed interval indicates the period when most participants were enrolled in the program. The earnings data come from the social security earnings SER records. Earnings include zero earnings for nonworkers.

The program's emphasis on job search might explain the temporary increase in earnings after participants leave it. Job Corps centers use placement outcomes in their evaluation of vocational instructors and other staff. Staff from Job Corps centers often remain in contact with former participants to assist in subsequent job searches.

9.2.2 National Guard ChalleNGe

The National Guard Youth ChalleNGe program is a 17-month intervention for youth who have dropped out of high school. Although the program does not require that participants enroll in the military, it stresses aspects of military discipline. Like the Job Corps, most training occurs in a residential facility removed from the usual environments of participants. The program features a two-week residential orientation and assessment period, a 20-week residential program often conducted at a military base, and a one-year, nonresidential mentoring program. There have been three major evaluations of the program: one occurring nine months after random assignment, another 21 months after random assignment, and a third evaluation three years after random assignment (Bloom et al., 2009; Millenky et al., 2010, 2011).

Table 9 shows the estimated treatment effects of the program. Nine months after the program, participants are more likely to have a GED or high school diploma and be employed. They are less likely to have been arrested or convicted. Three years after the intervention, the effects for criminal behavior and high school graduation decline and become statistically insignificant. The initial reduction in crime likely occurs because participants are housed in a residential program (an incapacitation effect). However, the employment and earnings effects persist. Given the similarities between the Job Corps and the ChalleNGe program, the three-year follow-up might not be long enough to determine whether the labor market effects are truly persistent.

In-depth interviews of the participants shed light on why some of the effects of the program might fade. Most participants were glad that they had entered the program, but explained that friends made it difficult to maintain the discipline that they had learned once they returned to their original environments (Millenky et al. (2011)). The authors suggest that it might be important to maintain stronger attachment with participants after they leave the residential facility.

Table 9 Summary of Treatment Effects from the National Guard ChalleNGe Program

	Time Since Assignment		
Outcome	9 Months$^{(1)}$	21 Months$^{(2)}$	3 Years$^{(3)}$
Has HS Diploma(%)	12.0***	5.7**	3.7
Has GED(%)	23.4***	26.5***	22.4***
Earned Any College Credit	–	15.1***	16.1***
Currently Working(%)	9.1***	4.9*	7.1**
Earnings (Annual)$^{(a)}$	–	2,028***	2,226***
Working or in Training(%)	−1.9	5.7**	4.6
Arrested(%) $^{(b)}$	−5.8**	0.9	−0.8
Convicted(%)$^{(b)}$	−4.4**	−4.2**	2.8
High Self-Efficacy	4.0**	–	–
Low Self-Efficacy	−9.6***	–	–
Serious Psychological Distress$^{(c)}$	–	0.3	−1.1
Always Uses Birth Control	–	0.3	−8.0**
Overweight	4.2	1.0	6.1**
Obese	−4.3**	−1.5	−0.7

Notes: The estimates are coefficients from regressions that control for sample member characteristics. The headers in the columns indicate the duration since random assignment. (a) For the 21-month survey, annual earnings are approximated by multiplying the estimates for weekly earnings by 52. Nonworkers are included as "0's." (b) For the 9-month survey, the value is for differences since random assignment. For the 21-month and 3-year follow-ups the value indicates differences for the past 12 months. (c) The psychological distress score is based on the K6 scale, which ranges from 0 to 24. Serious psychological distress is a score of 13 points or more.
* 10% significance; ** 5% significance; *** 1% significance.
Sources: (1) The estimates after 9 months come from Bloom et al. (2009). (2) The estimates after 21 months come from Millenky et al. (2010). (3) The estimates after 3 years come from Millenky et al. (2011).

9.3 Workplace-Based Adolescent Intervention Programs

Workplace-based internships and apprenticeships can boost skills.[104] A hundred and fifty years ago, apprenticeships and workplace-based education were standard for most adolescents. Apprenticeships offer in-depth, work-based learning combined with related academic course work. Few rigorous studies have examined how entering and completing apprenticeships in the United States affects the educational attainment, job skills, non-academic

[104]We have benefitted from the commentary of Robert Lerman in preparing this subsection. See Lerman (2013) for a discussion of related issues.

skills, and job market outcomes of young people.[105] A study of high school students who participated in a Wisconsin youth apprenticeship in printing documented participant earnings levels substantially above expected earnings for similar youth (Orr, Bloom, Bell, Lin, Cave, and Doolittle, 1994). Anecdotal evidence suggests that youth apprenticeships motivate participants to do better in school and to pursue difficult courses broadly related to their occupational interests.

Well-structured career-focused programs can potentially enhance character as well as occupational skills, in ways most relevant to the occupations and tasks that engage young people. Halpern (2009) undertook in-depth, observational studies of high school youth apprenticeships over a three-year period. He analyzed 24 separate programs involving nearly 500 apprentices and conducted over 300 hours of observation and over 90 interviews with adult mentors, staff, program directors, and students. One program, called After School Matters, engaged inner-city Chicago high school students in one- to two-year programs for two to three afternoons a week in fields such as video and computer technology, culinary arts, and design. Halpern also followed students in other programs, including Wisconsin's youth apprenticeship program.

Halpern's qualitative analysis of these programs is rich and difficult to summarize. Nonetheless, several perspectives are worth noting. He gives many examples of how apprentices gradually develop expertise in an occupation as well as in problem-solving skills. They acquire self-confidence, teamwork, the ability to take direction and take initiatives, and other noncognitive skills. He notes that participating youth see themselves as judged by their ability to meet or exceed the established standards of a discipline, including meeting deadlines and facing genuine constraints and unexpected difficulties that arise in the profession. To quote

[105]Students involved in Wisconsin's Youth Apprenticeship Program believed their worksite learning experiences were excellent, with 84%–86% reporting that they enabled them to improve their problem-solving and teamwork skills (Scribner and Wakelyn, 1998). Students reported gaps between their work-based and school-based learning. But interviews with instructors and employers indicated that apprentices improved their social and interpersonal skills, develop independent decision-making skills, and increased their self-confidence and self-esteem. However, one should be wary of self-reports. Students in the Cambridge-Somerville program reported satisfaction with the program, even though objective evaluations of it were negative (McCord, 1978).

Halpern, "Young people learn through observation, imitation, trial and error, and reiteration; in other words through force of experience." Adult mentors scaffold the discipline protocol for the apprentice, sequencing and controlling task demands to keep them on the constructive side of difficulty. They direct apprentices' attention, demonstrate new points, and sometimes collaborate with them.

Utilizing workplaces as learning locations is motivated by evidence of the importance of occupational and character skills. Workplace-oriented training can help youth attain development goals such as personal autonomy, efficacy, motivation, realism, optimism, and knowledge of vocations. It facilitates matching between workers and firms, and motivates adolescents to acquire relevant academic and character skills. Evidence from other countries shows that workplace-based learning helps students develop an occupational identity, a professional ethic, and self-esteem based on accomplishment (Rauner, 2007). We review some of the better documented programs in this section.

9.3.1 Career Academies

Career Academies provide evidence that integrating career development into a standard high school can have long-term labor market benefits. These academies operate within regular high schools and prepare students for postsecondary education and employment. In addition to regular high school courses, Career Academies offer courses that are career- or occupation-focused (Kemple and Willner, 2008). Individual academies focus on preparing students for participation in particular industries such as finance, health science, hospitality, and tourism. They include activities that might enhance character skills in the workplace (Kemple and Snipes, 2000). Career Academies have many features in common with European apprenticeship programs that are discussed in Section 9.10

Compared to other high schools, Career Academies expose students to career-oriented activities, including job shadowing, career fairs, job search workshops, resume preparation, and interview preparation. Students work together in teams when relating their experiences.

A recent evaluation based on random assignment estimates the effects of Career Academies on labor market, educational, and social outcomes (Kemple and Willner, 2008). Table 10 shows the difference in outcomes between the control and treatment groups. Career Academies have positive and sustained impacts on earnings and employment among young men but have no impact on high school completion rates and postsecondary educational enrollment and attainment rates.[106] Treatment group participants are more likely to live independently with children and a spouse/partner. Young males in the treatment group are more likely to get married and become custodial parents. For females, most effects are positive but not statistically significant.

Career Academies likely benefit their students by improving character skills. The program improves earnings for males in the long run, even though it does not improve educational attainment or scores on achievement tests. The Career Academies include activities that are designed to improve character. Internships may teach students the importance of perseverance and Conscientiousness, along with other occupation-related skills that improve their labor market readiness. Programs that integrate school and work not only motivate students to learn relevant academic material but also integrate adolescents into the larger society and teach children the skills valued in the workplace and in society at large.

[106]The percentage gains in earnings were highest for the students facing the highest risk of dropping out of school. Young women did not experience any statistically significant gains in earnings. Given that only 55% of the treatment group actually finished the full curriculum of the Career Academy, the earnings gains likely understate the actual impact of full participation.

Table 10 Summary of Treatment Effects from Career Academies within 96-Month Follow-Up after Scheduled High School Graduation

Outcome	**Males**	**Females**
Labor Market (49–96 Months)		
Monthly Earnings ($)$^{(a)}$	361**	118
Months Employed (#)	2.8***	−0.3
Average Hours Worked per Week (#)	4.1***	0.2
Average Hourly Wages ($)	0.6	0.7
Educational Attainment (After 96 Months)		
High School Diploma	−0.4	0.2
GED	3.6	1.3
Certificate/License	2.0	0.1
AA Degree	−1.0	1.8
BA Degree	−2.2	−1.6
Family Formation (After 96 Months)		
Married and Living Together	9.0**	1.5
Custodial Parent	11.5***	3.7
Non-Custodial Parent	−6.4**	0.2

Notes: Impact estimates are regression-adjusted to control for background characteristics of the sample and for the clustering of students within schools and random assignment years. (a) Nonworkers were assigned a value of "0" for monthly earnings.
*10% significance; **5% significance; ***1% significance.
Source: Kemple and Willner (2008).

9.3.2 Year-Up Program

The Year-Up program provides low-income, mostly black and Hispanic 18- to 24-year-olds with full-time classes and six-month internships. To enter the program, applicants must submit a written essay, letters of reference, a high school transcript, and attendance records (Grobe, Rosenblum, and Weissman, 2010). Given the requirements, Year-Up applicants may be more motivated than their peers.

During the first six months of the program, students take classes involving business writing, time management, teamwork, problem solving, and conflict resolution (Grobe et al. (2010)). To remain in the program, students must maintain high attendance rates, be on time, and complete assignments (Roder and Elliot, 2011). Students who complete classes

are placed in internships with companies (Roder and Elliot, 2011). Participants continue to attend a weekly class during their internships where they share experiences about their internships, work on their resumes, do mock job interviews, and look into postsecondary options. The technical skills training component of the program includes instruction in word processing, spreadsheet, and presentation software. Some students learn about computer installation and repair, and other students learn how to track portfolios.

Roder and Elliot (2011) study the effect of Year-Up on labor market outcomes one year after the end of the program. Qualified applicants on the waiting list were assigned to treatment or control groups. Table 11 summarizes the effects of the program two years after random assignment (one year after the end of the program). The treatment group has higher annual earnings, primarily due to their increased hourly wages. The employment rates are similar between the treatment and control groups, although the treatment group was more likely to have full-time jobs. The earnings and wage benefits arise from improvements in job quality. Compared to the control group, more members in the treatment group find higher paying jobs in information technology and investment operations. About a quarter of the treatment group is hired by the employer sponsoring their internship.

The one-year follow-up is too short to determine whether the program has lasting effects, but some evidence suggests that the benefits might be more permanent than other similar programs. The Job Corps primarily improves earnings in the short run by increasing employment (Schochet et al. (2008)). In contrast, Year-Up increases hourly wages (Roder and Elliot, 2011). If these wage increases reflect increased human capital formation, the benefits likely last longer. Many employers of interns later hired their trainees.

The evaluation has no direct measures of character skills, but it does provide clues that the program affected character in ways that lead to better outcomes in the job market. Year-Up students were taught that appropriate behavior, dress, and communication are important for success.

Table 11 Impacts of the Year-Up Program

Outcome	Treatment	Control	Treatment Effect
Hourly Wage at Current or Most Recent Job	12.58	10.32	2.26***
Annual Earnings during Year after Program[(a)]	15,082	11,621	3,461*
Employed during Year after Program	86	83	3
Works Full-Time	67	55	12*
Works in Information Technology	22	2	20†
Works in Investment Operations	15	0	15†

Notes:(*a*) The annual earnings calculations include nonworkers.
*10% significance; **5% significance; ***1% significance; † No measure of statistical significance reported.
Source: Roder and Elliot (2011).

9.3.3 Dominican Youth Employment Program

A recent study by Gertler and Martinez (2012) evaluates the Dominican Youth Employment program. The intervention targets unemployed youth (ages 16–29) without a high school diploma. In the evaluation, youth were assigned to one of three groups: a group that received vocational skills training, life skill training, and an internship; a group that received life skills training and an internship; or a control group. The vocational training is available in a variety of fields including cosmetology, sales, tourism, and electricity. The life skills training focuses on building self-esteem and capacities for communication, conflict resolution, time management, and teamwork. The internship takes place at a private company and lasts 240 hours.

Table 12 summarizes treatment effects from the program 18 months after assignment for the group that received both treatments ("Life+Vocation") and the group that only received life skills training ("Life Only"). Females in both treatment groups have higher employment rates, higher earnings, and fewer pregnancies than their control group counterparts. These differences are statistically significant when compared with the control group. The difference in impact between the two treatment groups is only statistically significant for preventing pregnancies. Since both treatment groups benefit equally, the program likely helps females

by improving character or through work experience rather than through the receipt of vocational training. For males, life skills training had no effect, while the combined training produced negative impacts. Given that the males who received vocational training report that they are actively looking for work and that they believe that their employment outcomes will improve, the authors of the evaluation speculate that vocational training might have raised the labor market prospects of male participants.

Table 12 Summary of Treatment Effects from the Dominican Youth Employment Program 18 Months after Assignment

	Males		**Females**	
Outcome	Life + Vocation	Life Only	Life + Vocation	Life Only
Worked during Last Week (%)	-8.3^{**}	-3.8	6.4^{**}	7.2^{***}
Searching for Job in Last Week (%)	6.8^{*}	5.3	-4.5	-1.6
Monthly Wage Income	676	-414	456^{**}	518^{**}
Expect Better Employment (%)$^{(a)}$	3.8^{**}	1.9	2.9^{**}	3.2^{**}
Number of Children	–	–	-0.13^{*}	-0.15^{**}

Notes: The reported estimated effects come from a linear probability model. The tests of significance assume that the errors are clustered at the course level. (a) The variable takes the value of 1 if the respondent reports that they expect their employment to be improved a year from now and a value of 0 if they expect it to be the same or worse.
* 10% significance; ** 5% significance; *** 1% significance.
Source: Gertler and Martinez (2012).

9.3.4 Self-Sufficiency Project

The Canadian Self-Sufficiency Project (SSP) provides direct evidence that working improves character. The program provided earnings supplements to single parents on welfare if they found full-time jobs and agreed to leave welfare. New enrollees were told that if they remained on welfare a year later, they would have the chance to participate in the SSP program. After a year, they were either assigned to a treatment group that would receive an earnings subsidy for three years if they worked full time or to a control group that would not.[107]

[107]See Michalopoulos, Tattrie, Miller, Robins, Morris, Gyarmati, Redcross, Foley, and Ford (2002) for a detailed description of the program.

Assessing the effect of employment on skills is complicated by the possibility of reverse causality. Just as employment might affect skills, skills might affect employment. The SSP program provides a way to study the effects of exogenous variation of employment status on skills. Gottschalk (2005) uses assignment status in the experiment as an instrumental variable for employment to estimate the effect of employment on Locus of Control (the extent to which people feel they have control over their lives).[108] After 36 months, those who received the subsidy were more likely to have an improved Locus of Control, suggesting that work can improve this character skill. This finding supports the earlier evidence that work-based education can improve character.

Other studies evaluate the labor market benefits of the program. The findings are mixed. Michalopoulos et al. (2002) evaluate the labor market benefits of the program. During the three-year eligibility period, SSP had significant positive impacts on employment, earnings, and reduced welfare use. However, the effect faded. Nevertheless, the children of participants who were in preschool at randomization performed better in school. At the 36-month follow-up, the children of mothers in the treatment group scored 10 percentage points higher on a test of math skills ($p < 0.05$)[109] and about 1/8 of a standard deviation higher on a vocabulary test, but this difference is not statistically significantly different from zero.[110]

Later studies investigate short- and long-term treatment effects for the take-up group (the members of the treatment group who qualified for the supplements). Zabel, Schwartz, and Donald (2010) find that the take-up group was 25% more likely to be employed after 52 months. Another study by Zabel, Schwartz, and Donald (2013) shows that, although the overall treatment effect on the wages of the take-up group is small, there is heterogeneity in the impact of SSP on wages between "incentivized" and "nonincentivized" take-up groups.[111]

[108] Locus of Control is related to Big Five Neuroticism. See Almlund et al. (2011) for a discussion.

[109] The report does not specify the nature of the math skills test.

[110] The Peabody Picture Vocabulary Test-Revised was used to measure vocabulary.

[111] "Incentivized take-up group"—individuals who were induced by the program to find full-time jobs during a 13-month eligibility window and who would not otherwise have worked full time. "Nonincentivized take-up group"—members of the program group who would leave welfare and work full-time regardless of the SSP program.

For the incentivized program group the absolute wage progression was 10.8%–11.5% compared to 1.9%–2.2% for the control group while for the nonincentivized take-up group, the absolute wage progression was 11.2%–12.2% compared to 8.2%–11.6% for the control group. Card and Hyslop (2009) find that SSP increased welfare participation in the first year after initial entry and lowered it over the following five years even after SSP payments ended, although the effect largely disappeared after seven years.

10 Apprenticeship Programs[112]

Apprentices are employees who have formal agreements with employers to carry out a recognized program of work-based and classroom learning. Apprenticeships typically begin in late adolescence or early adulthood. The training is highly structured, with a well-designed curriculum usually lasting three to four years. It is common for apprentices to spend three to four days per week at workplaces, acquiring knowledge and undertaking productive work involving gradually increasing levels of complexity. Classroom instruction in theoretical and general studies takes place at vocational schools one to two days per week. Apprentices participate in the production process, work with a trainer who is often a mentor, and ultimately gain sufficient occupational mastery to become certified by an external body (Wolter and Ryan, 2011). The emphasis is on occupational skills, but apprentices are likely to improve character and other generic employment skills as well, including those most relevant for their chosen occupation. The training is employer-led and thus sensitive to market demands and less costly to the government than full-time schooling. Career Academies and apprenticeship programs share many features in common. They train participants in the skills of specific occupations, supplement the training with general education, and provide guidance and mentoring for participants.

Apprenticeship training is common in many countries, reaching 55%–70% of youth in Austria, Germany, and Switzerland. The scale of apprenticeship varies widely across coun-

[112]The commentary and text of Robert Lerman was helpful in preparing this section. See Lerman (2013).

tries. Apprenticeships as a share of the total labor force are about 3.7%–3.9% in Australia and Germany, about 2.3%–2.5% in Canada and England, and 1.7% in France but only about 0.3% in the United States.[113]

Given the concern that apprenticeships train workers too narrowly and make apprentices less adaptable to a changing labor market, in evaluating apprenticeship programs it is important to follow those completing apprenticeships over long time horizons. At a minimum, an evaluation has to find an appropriate comparison group, estimate costs to firms, workers, and the government, and take account of the heterogeneity of occupations used for apprenticeships.

Several studies of apprenticeship training in European countries find high rates of returns to participants, often in the range of 10%–15%.[114] Bougheas and Georgellis (2004) find that about 70% of apprentices were in full-time employment immediately after training. Cooke (2003) shows that apprentices do not initially earn more than other workers, but experience greater wage growth.

One recent study of apprenticeship training analyzes small Austrian firms that eventually failed. It compares the wages of apprentices who started well before the firm failed (and thus who had a long period in apprenticeship training) with wages of apprentices who spent less time in training because the firm failed earlier in their apprenticeship period (Fersterer, Pischke, and Winter-Ebmer, 2008). Estimates suggest that for a three- to four-year apprenticeship, post-apprenticeship wages are 12%–16% higher than they would otherwise be. Since the worker's costs of participating in an apprenticeship are often minimal, the returns are high.

What about job and occupational mobility? Buechtemann, Schupp, and Soloff (1993)

[113]Apprenticeships have expanded rapidly in several countries, including Australia (rising from 157,000 in 1996 to 515,000 in 2012) and in England (rising from 53,000 in 1990 to 614,000 in 2012) (National Apprenticeship Service, 2012; National Centre for Vocational Education Research, 2012). Along with the increasing roles of apprenticeship in Australia and England have come initiatives to upgrade the quality of their programs (McDowell, Oliver, Persson, Fairbrother, Wetzlar, Buchanan, and Shipstone, 2011).

[114]See, for example, Adda, Dustmann, Meghir, and Robin (2013); Bougheas and Georgellis (2004); Clark and Fahr (2002); Winkelmann (1996).

find that about 80% of apprentices join the firm where they apprenticed after completing the program. In order to examine the adaptability of apprenticeship training, researchers (Clark and Fahr, 2002; Geel and Backes-Gellner, 2009) identified groups of skills that are required for each occupation. Changing occupations may not imply a significant change in the types of expertise required because of the overlap of skill requirements across occupations. While only 42% of apprentices stay in their initial occupation, nearly two-thirds remain with the apprentice occupation or another occupation requiring a similar mix of skills (Geel and Backes-Gellner, 2009). When apprentices move to another occupation that requires similar skills, their wages increase. Those trained in occupations with more specific skills are most likely to remain in their initial occupation or move to occupations within the same cluster.[115]

Germany has formally integrated apprenticeship programs into its educational system with apparent success. In Germany, there are three types of secondary schools: low (Hauptschule), medium (Realschule), and high (Gymnasium). Students who graduate from any of these secondary schools qualify to participate in the dual apprenticeship program. Those who graduate from higher quality schools have more prestigious vocational options.[116] The apprenticeships typically last about three years during which participants spend one to two days per week in a part-time vocational training school. They work the remaining time (Franz and Soskice, 1995).

Harhoff and Kane (1997) argue that apprentices occupy a similar place in the German economy as high school graduates occupy in the American economy. They find that apprentices earn about 19% more than those completing 10 or fewer years of school but 47% less than college graduates, whereas in the United States high school graduates earn about 23% more than those who complete 10 or fewer years of school but 42% less than those who complete college. The estimates are similar over the life cycle.

[115]Clark and Fahr (2002) also find that the wage penalty of departing from the occupation for which apprentices trained varies with the distance away from the original occupation and that there is no penalty at all from displacement into a somewhat related occupation. Among the former apprentices changing occupations, about two of five report using many or very many of the skills from their apprenticeship and another 20% used some of the skills.

[116]See Franz and Soskice (1995) for a detailed description of the program.

Adda et al. (2013) estimate that participating in the German apprenticeship program yields a 9%–10% annual rate of return. This estimate includes the return to character skills which the authors do not isolate from the other benefits of the program. It also includes the return to the education and job training received, which likely produces skills valued in the labor market.

Some American apprenticeships are also promising. Hollenbeck (2008) estimates rates of return to various types of job training in Washington State. He compares outcomes of apprentices to those of other workers with similar backgrounds and earnings histories. His results show that apprenticeship training yields higher returns than other training or postsecondary (mostly community college) occupational programs. For apprenticeships, the social and governmental rates of return are large (over 20%) within the first 2.5 years after apprentices exit the program.

Another, more recent study of apprenticeship in 10 American states documents large and statistically significant earnings gains from participating in an apprenticeship (Reed, Yung-Hsu Liu, Kleinman, Mastri, Reed, Sattar, and Ziegler (2012)). The study estimates how the length of participation in an apprenticeship affects earnings, holding constant the pre-enrollment earnings of apprenticeship participants. Six years after starting a program, the earnings of the average apprenticeship participant (average duration in an apprenticeship) are 40% higher than the earnings of nonparticipants with the same pre-apprenticeship history.

Workplace-based education is promising. It breaks down the artificial barrier between the culture of high schools and that of the rest of society documented in Coleman (1961). It motivates adolescents to learn relevant academic skills and gives mentoring and guidance to participants, many of whom come from disadvantaged families where such guidance is missing.

11 Other Curricula that Have Been Applied to Multiple Age Groups

11.1 Tools of the Mind

A random assignment evaluation of *Tools of the Mind* provides evidence that character skills can be fostered. *Tools of the Mind* is a program that attempts to teach preschool and early primary schoolchildren to regulate their social and cognitive behaviors. The curriculum encourages children to role-play and learn in groups with other children. In short-term follow-ups, several studies show that it improves classroom behavior as well as executive functioning, defined as higher-level cognitive skills, including inhibitory control, working memory, and cognitive flexibility.[117] Similar findings are reported for the Montessori preschool curriculum (Lillard and Else-Quest, 2006).

However, the evidence on the program is not all favorable.[118] There are few long-term evaluations of it. As noted by Heckman et al. (2013), the Perry curriculum and the *Tools of the Mind* curriculum have common intellectual origins, so the positive long-term evidence on Perry arguably transfers to forecasting the likely long-run effects of *Tools of the Mind*.

11.2 Studies that Teach the Incremental Theory of Intelligence

Several studies examine the effect of teaching students that capabilities are malleable and that learning can change the structure of the brain (the "incremental theory of intelligence"). Dweck (2007) presents an extensive discussion of this approach. The focus in this intervention is on promoting cognition, but the measures used to evaluate the interventions include achievement tests and grades that are also determined by character skills. Some of this research also measures perseverance. The theory underlying the intervention is that children

[117] Barnett, Jung, Yarosz, Thomas, Hornbeck, Stechuk, and Burns (2008); Barnett, Yarosz, Thomas, and Hornbeck (2006); Bodrova and Leong (2001, 2007); Diamond, Barnett, Thomas, and Munro (2007).

[118] See the study by Farran, Lipsey, and Wilson (2011).

who believe that cognitive ability is a fixed trait might have little incentive to improve it.

The "mindset" interventions implement the logic of Figure 1. Effort and character skills determine performance on tasks, including performance on grades and achievement tests. Motivating greater effort enhances performance, and the "mindset" intervention is one such motivational exercise. At issue is whether the enhanced effort, however motivated, has lasting consequences. Unfortunately, all of the follow-ups are short-term in nature, so the important question—do these motivational programs have lasting consequences?—remains unanswered.[119]

Table 13 summarizes the results from three evaluations of these programs. Aronson, Fried, and Good (2002) evaluate an intervention in which college students were taught the incremental theory of intelligence. As a way to internalize the information, the participants were instructed to write a letter that explains the theory to a low-performing middle school student. At the end of the next academic quarter, the average GPA in the treatment group was modestly higher than in the control group. There was no follow-up.

[119]Motivation research (Deci, 1971; Deci and Ryan, 1985) suggests that intrinsic rewards have longer lasting consequences than extrinsic rewards (e.g., payments or punishment).

Table 13 Evidence on the Treatment Effects of Teaching that Intelligence is Malleable

Intervention	Follow-up	Main Results	Source
College students were taught the incremental theory of intelligence and wrote letters to a middle school student explaining the theory on three occasions over a one month period.	2 quarters[(a)]	0.20–0.27 point increase in GPA ($p < 0.05$)[(b)]	Aronson et al. (2002)
Seventh graders received explanations of the incremental theory of intelligence via two 90-minute, in-person meetings with a mentor and weekly emails during the school year.	1 year	0.64 standard deviation increase on math score for boys ($p < 0.10$), 1.13 standard deviation increase in TAAS math scores for females ($p < 0.01$), and 0.52 standard deviation increase in reading scores ($p < 0.05$)[(c)]	Good, Aronson, and Inzlicht (2003)
Seventh graders attended a weekly, 25-minute class that explained the incremental theory of intelligence.	1 semester	0.53 point increase in math GPA ($p < 0.05$)	Blackwell, Trzesniewski, and Dweck (2007)

Notes: (a) The intervention began at the beginning of the winter quarter, and grades were measured at the end of the spring quarter. (b) A range is given because the effect size differs across races and between different types of comparison groups. Not all of the effects are significant at the 5% level. (c) The authors do not estimate separate effects for males and females on the reading score.

Good et al. (2003) study an intervention in which seventh-grade students were taught the incremental theory of intelligence. At the end of the year, students who learned about the incremental theory of intelligence performed somewhat better on the math and reading sections of the Texas Assessment of Academic Skills (TAAS) achievement test. The effect was strongest for girls.

Blackwell et al. (2007) evaluate a program in which seventh graders are taught about the incremental theory of intelligence through a series of class sessions. The treatment group and control group participated in eight 25-minute classroom sessions that had similar course material about the physiology of the brain, study skills, and antistereotypic thinking. Only the treatment group was taught the incremental theory of intelligence. After one term, the treatment group improved its math GPAs relatively more than the control group.

While these studies are promising, their follow-ups are very short. The longest follow-up is a year after the beginning of the program. For example, it is possible that the interventions increased motivation in the short run, much like the studies in which children are incentivized using M&M candies. In addition, while the effects are statistically significant, they are relatively small. It is unknown whether the benefits outweigh the costs. None of the studies of the "mindset" intervention report a rate of return. The mechanism by which reported results are produced is not analyzed.[120]

12 The Effects of Education and Parental Investment on Character and Cognition

Studies that analyze the consequences of parental and schooling investment demonstrate their causal effects on promoting cognitive and character skills. Heckman, Stixrud, and Urzúa (2006) estimate the effect of schooling in creating cognitive and character skills, controlling

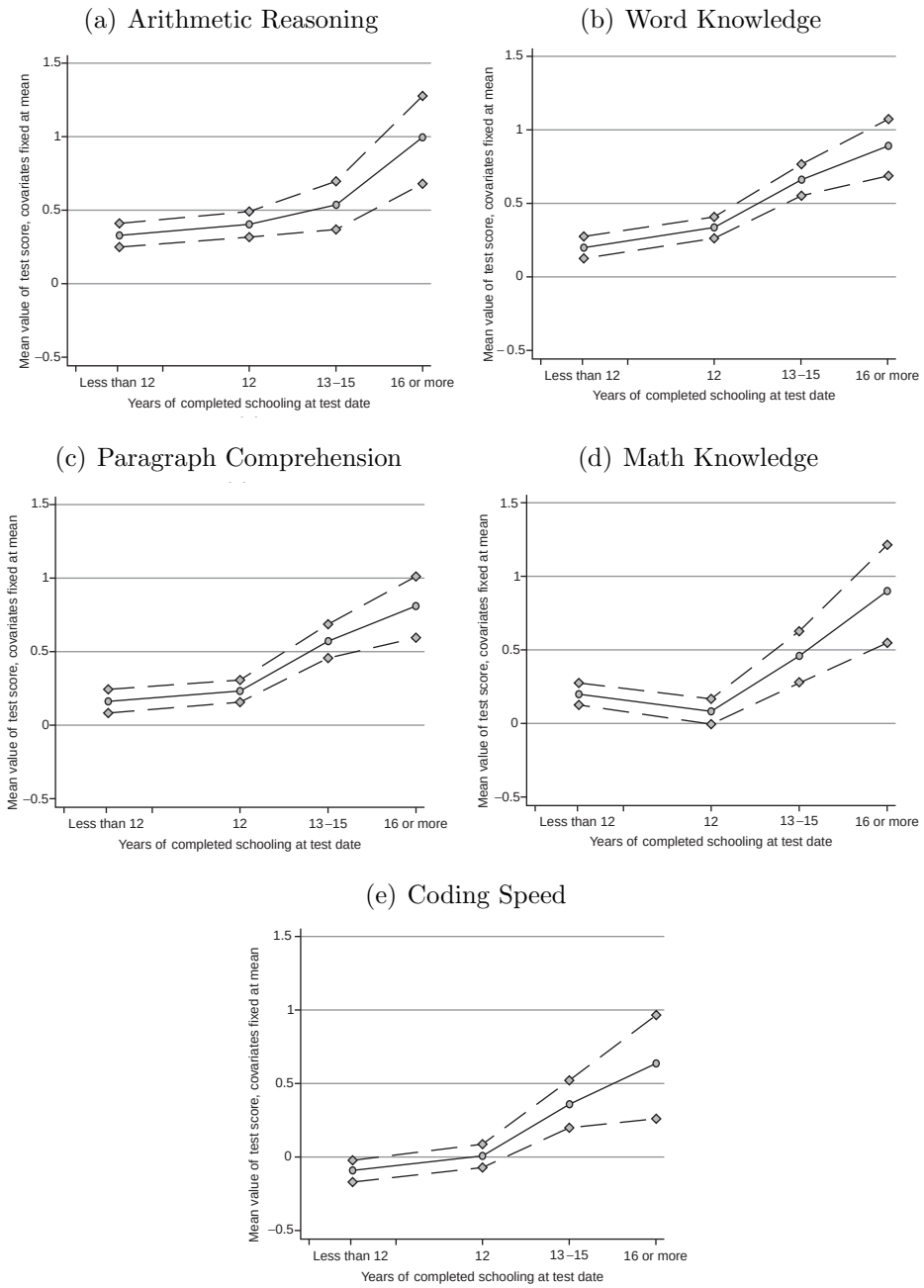
[120]Hunt (2012), a leading authority on intelligence, provides a trenchant assessment of this group of interventions and in particular the work of Dweck.

for the problem of reverse causality that schooling may be caused by skills. Specifically, they estimate the effect of schooling on Self-Esteem and Locus of Control, character skills related to Neuroticism.[121] Figures 14 and 15 show the effect of education on measures of these skills: Schooling improves both character and cognitive skills. In results that are not displayed, they show that these traits, in turn, boost a variety of labor market and social outcomes at age 30.[122]

[121]The Rosenberg Self-Esteem Scale assesses the degree of approval or disapproval of oneself (Rosenberg, 1965). The relationship between these measures and the Big Five traits of Neuroticism is discussed in Almlund et al. (2011). See Table 1.

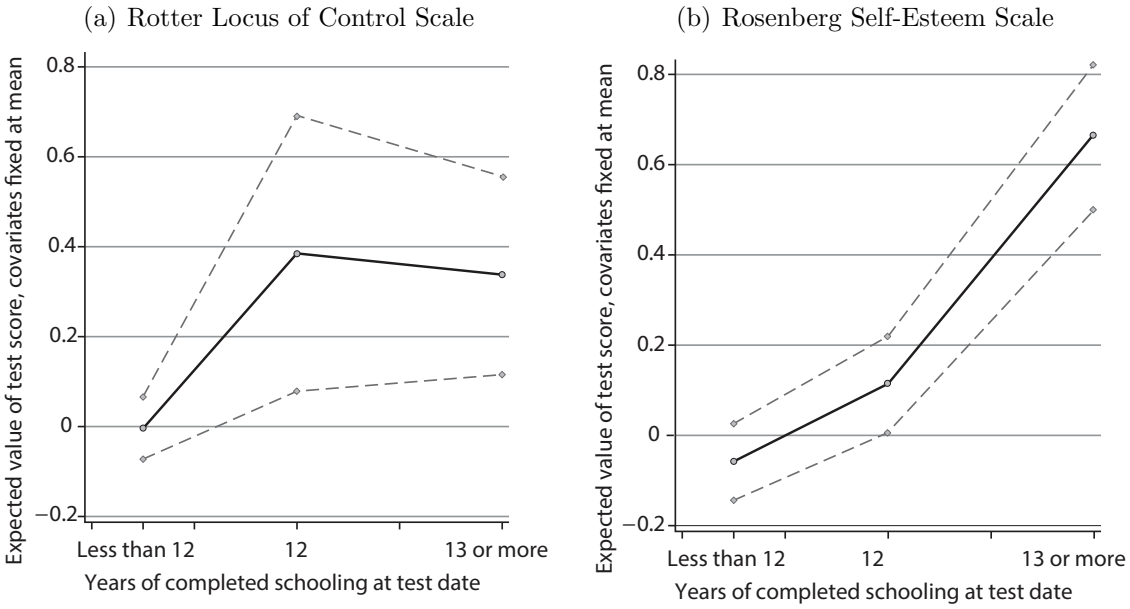
[122]Both Heckman, Humphries, Urzúa, and Veramendi (2011) and Heckman, Stixrud, and Urzúa (2006) use an identification strategy employing matching on proxies for unobserved skills that corrects for measurement error and the endogeneity of schooling.

Figure 14 Causal Effect of Schooling on ASVAB Measures of Cognition



Source: Heckman et al. (2006, Figure 4).
Notes: Mean effect of schooling on components of the ASVAB. The first four components are averaged to create males with average ability. We standardize the test scores to have within-sample mean zero and variance one. The model is estimated using the NLSY79 sample. Solid lines depict average test scores, and dashed lines, 2.5%–97.5% confidence intervals. Regressors are fixed at means.

Figure 15 Causal Effect of Schooling on Two Measures of Character



Source: Heckman et al. (2006, Figure 5).
Notes: Effect of schooling on socioemotional scales for males with average ability, with 95% confidence bands. The locus of control scale is based on the four-item abbreviated version of the Rotter Internal-External Locus of Control Scale. This scale is designed to measure the extent to which individuals believe they have control over their lives through self-motivation or self-determination (internal control), as opposed to the extent to which individuals believe that the environment controls their lives (external control). The self-esteem scale is based on the 10-item Rosenberg Self-Esteem Scale. This scale describes a degree of approval or disapproval toward oneself. In both cases, we standardize the test scores to have within-sample mean zero and variance one, after taking averages over the respective sets of scales. The model is estimated using the NLSY79 sample. Solid lines depict average test scores, and dashed lines, 2.5%–97.5% confidence intervals. Regressors are fixed at means.

Cunha, Heckman, and Schennach (2010) estimate a causal model of parental investment using longitudinal data on the development of children with rich measures of parental investment and child skills. It estimates the model of Figure 9. They show that skills are self-productive and exhibit dynamic complementarity; levels of skills at one age affect the productivity of future investments at later ages and hence help determine the evolution of future skills through direct and cross effects. For example, more motivated children learn more.

The authors find that it is more difficult to compensate for the effects of adverse environments on cognitive skills in adolescence than it is to build cognitive skills at earlier ages. Thus the early years are sensitive periods for cognitive skills. This finding is consistent with the high-rank stability of cognitive skills after age 11 or so; it also helps to explain the evidence on ineffective cognitive remediation strategies for disadvantaged adolescents documented in Cunha, Heckman, Lochner, and Masterov (2006), Knudsen, Heckman, Cameron, and Shonkoff (2006), and Cunha and Heckman (2007).

Character skills foster cognitive development. Greater malleability of character skills is found over longer stretches of the life cycle than for cognitive skills. This occurs in part because new aspects of character emerge with maturity and can be influenced.[123] Investment in character skills in the early years has a higher economic return than investment in the later years because it builds the base for subsequent investment. Nonetheless, the productivity of later-age investment in character skills is substantial. If the early years have been compromised, it is more effective in the adolescent years to focus on developing character skills rather than on cognitive skills.[124]

Heckman et al. (2011) estimate a sequential model of education to study the effects of education on a variety of outcomes, controlling for cognitive and character skills and the endogeneity of education. Correcting for selection into education, they find that early cogni-

[123]Borghans et al. (2008).

[124]Cunha, Heckman, and Schennach (2010) report that 16% of the variation in educational attainment is explained by adolescent cognitive skills, 12% is due to adolescent socioemotional traits (character skills), and 15% is due to measured parental investments.

tive and character skills have substantial causal effects on schooling, labor market outcomes, adult health, and social outcomes and that education at most levels causally produces benefits in labor market, health, and social outcomes. They also study the effects of schooling on measures of adult character and find substantial impacts.

Jackson (2013) estimates a valued-added model of the effect of ninth-grade math and English teachers on student cognitive and noncognitive (character) skills as measured by absences, suspensions, grades, and grade progression.[125] Similar models have been adopted by school districts to assess the teacher's impact on student achievement test scores and to determine teacher bonuses. Jackson estimates a bigger effect of teachers on character skills than on cognitive skills.[126] Estimated teacher effects on cognition and character are only weakly correlated with each other, suggesting that some teachers improve one dimension of skill without improving the other. These findings imply that using achievement tests alone to assess teacher effectiveness would miss important dimensions of teacher quality.

Coleman, Hoffer, and Kilgore (1982), Neal (1997), and Bryk, Lee, and Holland (1993) document the benefits of Catholic education in terms of scores on achievement tests.[127] It has long been speculated that Catholic schools mold character through creating a more disciplined environment and through teaching moral values. However, to our knowledge, there are no direct estimates of the effects of religious education on character. Typical of the entire literature on evaluation of schools, these studies focus on the effects of Catholic schools on test scores and schooling.

Arum (2005) documents the decline of disciplinary practices in American schools arising in part from legal and legislative developments originating in the 1960s that increased the rights of students and circumscribed teachers from using traditional measures of enforcing discipline. He cites evidence showing that greater discipline improves test scores and behavior

[125]See, for example, the Teacher Advancement Program in Chicago Public Schools (Glazerman and Seifullah, 2012).

[126]The estimates are from teacher "fixed-effect" models.

[127]Neal (1997) corrects for the selection effect that Catholic schools can expel unruly and undisciplined students. Uncontrolled comparisons could show a spurious positive effect of Catholic schools that do not reflect their true effect on any given person.

and promotes greater high school graduation.[128]

Some modern schools have shown how it is possible to integrate character education into schools. The Knowledge Is Power Program (KIPP) is a group of public charter middle schools that are designed to improve educational outcomes for low-income families. Tough (2012) discusses how the first wave of KIPP students excelled at taking achievement tests but dropped out of college at disappointingly high rates. In response, KIPP schools started to emphasize character skills, "grit, self-control, social intelligence (including self-advocacy), zest, optimism, and gratitude, that enable students to stick with college even in the face of considerable obstacles" (The KIPP Foundation, 2011). Their motto is: *Work hard. Be nice.*

Evidence on the efficacy of KIPP is lacking. An evaluation of KIPP schools suggests that it has improved achievement test scores three years after students enter the program.[129] A recent report from KIPP found that 33% of the students graduate from a four-year college, which is less than their goal of 75% but is four times higher than the rate for students with comparable backgrounds (The KIPP Foundation, 2011). However, these evaluations are based on comparing KIPP students to students in other schools; some have argued that KIPP students have higher levels of starting achievement and that KIPP selects more motivated students and parents into their schools.[130] A recent randomized trial evaluation of KIPP (Mathematica, 2013) found that KIPP boosted achievement test scores but had mixed effects on character skills. For example, participants in KIPP were more likely to lie to their parents or quarrel with them (Tuttle et al., 2013).

Similar to KIPP, a recent intervention in Chicago attempts to foster character skills. The OneGoal program selects and trains high school teachers to help students apply to colleges, improve grades and test scores, and, importantly, *persist* through college. It serves low-income schools in Chicago, most of which have college enrollment rates of less than 50%. Like KIPP, it attempts to cultivate character skills (Tough, 2012). So far, OneGoal

[128]DiPrete, Muller, and Shaeffer (1981).
[129]Tuttle, Teh, Nichols-Barrer, Gill, and Gleason (2010).
[130]Carnoy, Jacobsen, Mishel, and Rothstein (2005).

appears to be successful: 94% of OneGoal participants attend college, and 85% of the college enrollees complete the first year.[131] While there have been no long-term evaluations of KIPP or OneGoal that account for selection bias, these programs show that it is possible to openly emphasize character in schools without infringing on the rights of students or families or violating the boundaries between church and state. These efforts can be viewed as secular versions of moral and character education that do not introduce religion into schools.

13 Summary

This paper reviews recent literature on the economics and psychology of character skills and how interventions can develop character. Character skills predict later-life outcomes with the same, or greater, strength as measures of cognition.

Character is a skill—not a trait. It can be enhanced, and there are proven and effective ways to do so. Character is shaped by families and social environments. At any age, character skills are stable across different tasks, but performance on any task depends on multiple skills as well as the effort expended on it. Effort, in turn, depends on the incentives offered to perform the task. Since all measures of character and cognition are measures of performance on some task, it is necessary to standardize for incentives, effort, and other skills in measuring any particular character or cognitive skill. Despite these difficulties, reliable measures of character have been developed, although there is always room for improvement.

Though stable at any age, skills are not set in stone over the life cycle. Both cognitive and character skills can change. Parents, schools, and social environments shape them, although there are important genetic influences. Skill development is a dynamic process. The early years are important in laying the foundation for successful investment in the later years. While there is hard evidence on the importance of the early years in shaping all skills, some character skills are more malleable than cognitive skills at later ages.

[131] See `http://www.onegoalgraduation.org/onegoal-results/`. The OneGoal students are not old enough to estimate college graduation rates.

This paper reviews a variety of interventions targeted to different stages of the life cycle. Long-term evaluations of early childhood programs are more numerous. There is substantial evidence that high-quality early childhood programs have lasting and beneficial effects on character skills. The evidence on interventions in elementary schools shows lasting benefits of interventions that primarily operate through boosting character skills.

There are few long-term evaluations of adolescent interventions. The available evidence suggests a much greater benefit from programs that target character skills compared to the benefits of programs that mainly target cognition and academic learning. Workplace-based programs that teach character skills appear to be the most effective remediation interventions for adolescents. They motivate acquisition of work-relevant skills and provide discipline and guidance for disadvantaged youth that is often missing in their homes or high schools. Successful interventions at any age emulate the mentoring and attachment that successful families give their children.

The available evidence suggests that the most successful remediation programs are not as effective as the most successful early childhood programs. Building an early base of skills that promote later-life learning and engagement in school and society is a better strategy. Prevention is more effective than remediation.

References

Abbott, R. D., J. O'Donnell, J. D. Hawkins, K. G. Hill, R. Kosterman, and R. F. Catalano (1998). Changing teaching practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry 68*(4), 542–552.

Achenbach, T. M. and L. A. Rescorla (2001). *Manual for ASEBA School-Age Forms and Profiles*. Burlington, VT: Research Center for Children, Youth and Families.

Ackerman, P. L. and E. D. Heggestad (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin 121*, 219–245.

Adda, J., C. Dustmann, C. Meghir, and J.-M. Robin (2013). Career progression, economic downturns, and skills. Unpublished manuscript, Institute for Fiscal Studies, London, UK.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Amsterdam: Elsevier.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4 ed.). Washington, DC: American Psychiatric Association.

Aronson, J., C. B. Fried, and C. Good (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology 38*(2), 113–125.

Arum, R. (2005). *Judging School Discipline: The Crisis of Moral Authority*. Cambridge, MA: Harvard University Press.

Ayllon, T. and K. Kelly (1972). Effects of reinforcement on standardized test performance. *Journal of Applied Behavior Analysis 5*(4), 477–484.

Barnett, W. S., K. Jung, D. J. Yarosz, J. Thomas, A. Hornbeck, R. Stechuk, and S. Burns (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly 23*(3), 299–313.

Barnett, W. S., D. J. Yarosz, J. Thomas, and A. Hornbeck (2006). Educational effectiveness of a Vygotskian approach to preschool education: A randomized trial. Technical report, National Institute for Early Education Research, Rutgers, The State University of New Jersey.

Barrick, M. R. and M. K. Mount (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology 44*(1), 1–26.

Barton, P. (2006). *High School Reform and Work: Facing Labor Market Realities*. Princeton, NJ: Educational Testing Service.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012, July). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics 4*, 453–478.

Benda, B. B. (2005). The robustness of self-control in relation to form of delinquency. *Youth & Society 36*(4), 418–444.

Besharov, D. J., P. Germanis, C. A. Higney, and D. M. Call (2011). Houston Parent-Child Development Center. Working Paper 17, University of Maryland, School of Public Policy, Welfare Reform Academy.

Binet, A. and T. Simon (1916). *The Development of Intelligence in Children (The Binet-Simon Scale)*. Psychological Science. Baltimore, MD: Williams & Wilkins Co.

Blackwell, L. S., K. H. Trzesniewski, and C. S. Dweck (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development 78*(1), 246–263.

Bloom, D., A. Gardenhire-Crooks, and C. L. Mandsager (2009). Reengaging high school dropouts: Early results of the National Guard Youth ChalleNGe program evaluation. Report, MDRC. Last accessed online February 11, 2013. `http://www.mdrc.org/sites/default/files/full_491.pdf`.

Bodrova, E. and D. J. Leong (2001). *Tools of the Mind: A Case Study of Implementing the Vygotskian Approach in American Early Childhood and Primary Classrooms*. Geneva: International Bureau of Education, UNESCO.

Bodrova, E. and D. J. Leong (2007). *Tools of the Mind: The Vygotskian Approach to Early Childhood Education*. Upper Saddle River, NJ: Pearson Education, Inc.

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

Borghans, L., B. H. Golsteyn, J. J. Heckman, and H. Meijers (2009, April). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association 7*(2–3), 649–658.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2011a). Identification problems in personality psychology. *Personality and Individual Differences 51*(3: Special Issue on Personality and Economics), 315–320.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2011b). Reinterpreting estimated effects of cognition on social outcomes. Unpublished manuscript, Department of Economics, University of Chicago.

Borghans, L., H. Meijers, and B. ter Weel (2008, January). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry 46*(1), 2–12.

Bouchard, T. J. and J. C. Loehlin (2001, May). Genes, evolution and personality. *Behavior Genetics 31*(3), 243–273.

Bougheas, S. and Y. Georgellis (2004). Early career mobility and earnings profiles of German apprentices: Theory and empirical evidence. *Labour 18*(2), 233–263.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbooks in Economics*, Chapter 59, pp. 3705–3843. Amsterdam: Elsevier Science.

Bowen, W. G., M. M. Chingos, and M. S. McPherson (2009). Test scores and high school grades as predictors. In *Crossing the Finish Line: Completing College at America's Public Universities*, pp. 112–133. Princeton, NJ: Princeton University Press.

Bowlby, J. (1951). Maternal Care and Mental Health. Bulletin, World Health Organization, Geneva, Switzerland.

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

Bowles, S., H. Gintis, and M. Osborne (2001, December). The determinants of earnings: A behavioral approach. *Journal of Economic Literature 39*(4), 1137–1176.

Boyle, S. H., R. B. Williams, D. B. Mark, B. H. Brummett, I. C. Siegler, and J. C. Barefoot (2005). Hostility, age, and mortality in a sample of cardiac patients. *American Journal of Cardiology 96*(1), 64–66.

Breitmayer, B. J. and C. T. Ramey (1986, October). Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development. *Child Development 57*(5), 1151–1165.

Breuning, S. E. and W. F. Zella (1978). Effects of individualized incentives on norm-referenced IQ test performance of high school students in special education classes. *Journal of School Psychology 16*(3), 220–226.

Bridgeman, B., J. B. Blumenthal, and S. R. Andres (1981). Parent Child Development Center: Final evaluation report. Report, ETS Research.

Bryk, A. S., V. E. Lee, and P. B. Holland (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Buechtemann, C. F., J. Schupp, and D. J. Soloff (1993). Roads to work: school-to-work transition patterns in Germany and the United States. *Industrial Relations Journal 24*(2), 97–111.

Campbell, F., G. Conti, J. Heckman, S. Moon, and R. Pinto (2013). The long-term health effects of early childhood interventions. Under review, *Economic Journal*.

Campbell, F. A., E. P. Pungello, S. Miller-Johnson, M. Burchinal, and C. T. Ramey (2001, March). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology 37*(2), 231–242.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science 6*(1), 42–57.

Canli, T. (2006). *Biology of Personality and Individual Differences*. New York: Guilford Press.

Card, D. and D. R. Hyslop (2009). The dynamic effects of an earnings subsidy for long-term welfare recipients: Evidence from the Self-Sufficiency Project applicant experiment. *Journal of Econometrics 153*(1), 1–20.

Carneiro, P. and R. Ginja (2011, January). Preventing behavior problems in childhood and adolescence: Evidence from Head Start. Unpublished manuscript, University College London.

Carneiro, P. and R. Ginja (2012, January). Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start. IZA Discussion Paper 6315, Institute for the Study of Labor (IZA).

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?*, pp. 77–239. Cambridge, MA: MIT Press.

Carnoy, M., R. Jacobsen, L. Mishel, and R. Rothstein (2005). *The Charter School Dust-up.* Washington, DC: Economic Policy Institute.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* New York: Cambridge University Press.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, S. W. Diane, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics 126*(4), 1593–1660.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011, May). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics 126*(2), 749–804.

Clark, D. and R. Fahr (2002, February). The promise of workplace training for non-college bound youth: Theory and evidence from German apprenticeship. CEP Discussion Paper 518, Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Clingman, J. and R. L. Fowler (1976). The effects of primary reward on the I.Q. performance of grade-school children as a function of initial I.Q. level. *Journal of Applied Behavior Analysis 9*(1), 19–23.

Cloninger, C. R., D. M. Svrakic, C. Bayon, and T. R. Przybeck (1999). Measurement of psychopathology as variants of personality. In C. R. Cloninger (Ed.), *Personality and Psychopathology*, pp. 33–66. Washington, DC: American Psychiatric Press.

Coleman, J. S. (1961). *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education*. New York: The Free Press of Glencoe.

Coleman, J. S., T. Hoffer, and S. Kilgore (1982). *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.

Cooke, L. (2003). A comparison of initial and early life course earnings of the German secondary education and training system. *Economics of Education Review 22*(1), 79–88.

Costa, P. T. and R. R. McCrae (1992a). Four ways five factors are basic. *Personality and Individual Difference 13*(6), 653–665.

Costa, P. T. and R. R. McCrae (1992b). *Revised NEO Personality Inventory (NEO PI-R) and the NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

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

Cunha, F. and J. J. Heckman (2009, April). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*(2–3), 320–364. Presented as the Marshall Lecture, European Economics Association, Milan, Italy, August 29, 2008.

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

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Currie, J. and D. Thomas (1995, June). Does Head Start make a difference? *American Economic Review 85*(3), 341–364.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology 18*, 105–115.

Deci, E. L. and R. M. Ryan (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

DeLong, J. and K. Magin (2009, Winter). The U.S. equity return premium: Past, present and future. *Journal of Economic Perspectives 23*(1), 193–208.

Deming, D. (2009, July). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics 1*(3), 111–134.

Devlin, B., M. Daniels, and K. Roeder (1997). The heritability of IQ. *Nature 388*(6641), 468–471.

DeYoung, C. G., J. B. Hirsh, M. S. Shane, X. Papademetris, N. Rajeevan, and J. R. Gray (2010). Testing predictions from personality neuroscience: Brain structure and the Big Five. *Psychological Science 21*(6), 820–828.

Diamond, A., S. Barnett, J. Thomas, and S. Munro (2007). Preschool program improves cognitive control. *Science 318*(5855), 1387–1388.

DiPrete, T. A., C. Muller, and N. Shaeffer (1981). *Discipline and order in American high schools*. Washington, DC: National Center for Education Statistics. Contractor Report from NORC.

Duckworth, A. (2012, November). Reference bias. Presented at the University of Chicago.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007, June). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.

Duckworth, A. L., P. D. Quinn, D. R. Lynam, R. Loeber, and M. Stouthamer-Loeber (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences 108*(19), 7716–7720.

Durlak, J. A., R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development 82*(1), 405–432.

Dweck, C. S. (2007). *Mindset: The New Psychology of Success*. New York, NY: Ballantine Books.

Eckenrode, J., M. Campa, D. W. Luckey, C. R. Henderson, R. Cole, H. Kitzman, E. Anson, K. Sidora-Arcoleo, and D. L. Olds (2010, January). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Journal of the American Medical Association 164*(1), 9–15.

Edlund, C. V. (1972). The effect on the behavior of children, as reflected in the IQ scores, when reinforced after each correct response. *Journal of Applied Behavior Analysis 5*(3), 317–319.

Engle, P. L., M. M. Black, J. R. Behrman, M. Cabral de Mello, P. J. Gertler, L. Kapiriri, R. Martorell, M. Eming Young, and The International Child Development Steering Group

(2007, January). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet 369*(9557), 229–242.

Engle, P. L., L. C. H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011, October). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet 378*(9799), 1339–1353.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology 37*(7), 1097–1126.

Farran, D. C., M. W. Lipsey, and S. Wilson (2011, August). Experimental evaluation of the Tools of the Mind pre-k curriculum. Technical report, Peabody Research Institute Report.

Fersterer, J., J.-S. Pischke, and R. Winter-Ebmer (2008). Returns to apprenticeship training in Austria: Evidence from failed firms. *Scandinavian Journal of Economics 110*(4), 733–753.

Flynn, J. R. (2007). *What is Intelligence?: Beyond the Flynn Effect*. New York: Cambridge University Press.

Franz, W. and D. W. Soskice (1995). The German apprenticeship system. In *Institutional Frameworks and Labor Market Performance*, Chapter 8, pp. 208–234. New York, NY: Routle.

Friedman, D. and A. H. Streicher (1985). Reliability of scores for Fiscal Year 1981 Army applicants: Armed Services Vocational Aptitude Battery forms 8, 9, and 10. Technical report, Defense Technical Information Center.

Friedman, H. S. and L. R. Martin (2011). *The Longevity Project: Surprising Discoveries for Health and Long Life from the Landmark Eight-Decade Study*. New York, NY: Hudson Street Press.

Garces, E., D. Thomas, and J. Currie (2002, September). Longer-term effects of Head Start. *American Economic Review 92*(4), 999–1012.

Geel, R. and U. Backes-Gellner (2009). Occupational mobility within and between skill clusters: An empirical analysis based on the skill-weights approach. Economics of Education Working Paper Series 0047, University of Zurich, Institute for Strategy and Business Economics (ISU).

Gensowski, M. (2012). Personality, IQ, and lifetime earnings. Unpublished manuscript, University of Chicago, Department of Economics.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, and S. Grantham-McGregor (2013). Labor market returns to early childhood stimulation: A 20-year followup to an experimental intervention in Jamaica. Working Paper 19185, NBER.

Gertler, P. and S. Martinez (2012). Vocational and life skills in youth training: A randomized experiment in the Dominican Republic. Presented at the IZA/World Bank/OECD Conference on Activation and Employment Support Policies, Bahcesehir University, Istanbul, Turkey, May 01, 2012.

Glazerman, S. and A. Seifullah (2012). An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years. Final Report 06736-520, Mathematica Policy Research.

Goldammer, C. (2010). Skills, investment, and race: Evidence from the ECLS-K. Unpublished manuscript, University of Chicago, Department of Economics.

Goldin, C. and L. F. Katz (2008). *The Race between Education and Technology*. Cambridge, MA: Belknap Press of Harvard University Press.

Goldschmidt, P. and D. Huang (2007). The long-term effects of after-school programming on educational adjustment and juvenile crime: A study of the LA's BEST after-school program. Technical report, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles, CA.

Good, C., J. Aronson, and M. Inzlicht (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology 24*(6), 645–662.

Gottfredson, M. R. and T. Hirschi (1990). *A General Theory of Crime*. Palo Alto, CA: Stanford University Press.

Gottschalk, P. (2005). Can work alter welfare recipients' beliefs? *Journal of Policy Analysis and Management 24*(3), 485–498.

Gough, H. G. and A. B. Heilbrun (1983). *The Adjective Check List Manual*. Palo Alto, CA: Consulting Psychologists Press.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet 369*(9555), 60–70.

Grantham-McGregor, S. M., C. A. Powell, S. P. Walker, and J. H. Himes (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. *The Lancet 338*(8758), 1–5.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy 24*(1), 23–37.

Griffin, M., J. J. Heckman, and S. H. Moon (2013). Preschool curricula. Unpublished manuscript, University of Chicago.

Griffiths, R. (1954). *The Abilities of Babies*. London, UK: University of London Press.

Grobe, T., E. Rosenblum, and T. Weissman (2010). *Dollars and Sense: How "Career First" Programs Like Year Up Benefit Youth and Employers*. Boston, MA: Jobs for the Future.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics 19*(3), 403–420.

Haggerty, K., R. Kosterman, R. F. Catalano, and J. D. Hawkins (1999). Preparing for the drug free years. Juvenile justice bulletin, US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

Halpern, R. (2009). *The Means to Grow Up: Reinventing Apprenticeship as a Developmental Support in Adolescence* (1 ed.). Critical youth studies. New York, NY: Routledge.

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. (2003, February). The failure of input-based schooling policies. *Economic Journal 113*(485), F64–F98.

Harhoff, D. and T. J. Kane (1997). Is the German apprenticeship system a panacea for the US labor market? *Journal of population economics 10*(2), 171–196.

Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver, CO: University of Denver.

Hawkins, J., E. Von Cleve, and R. Catalano Jr (1991). Reducing early childhood aggression: Results of a primary prevention program. *Journal of the American Academy of Child & Adolescent Psychiatry 30*(2), 208–217.

Hawkins, J. D., R. F. Catalano, R. Kosterman, R. D. Abbott, and K. G. Hill (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics and Adolescent Medicine 153*(3), 226–234.

Hawkins, J. D., R. Kosterman, R. F. Catalano, K. G. Hill, and R. D. Abbott (2005). Promoting positive adult functioning through social development intervention in childhood. *Archives of Pediatrics and Adolescent Medicine 159*(1), 25–31.

Hawkins, J. D., R. Kosterman, R. F. Catalano, K. G. Hill, and R. D. Abbott (2008). Effect of social development intervention in childhood 15 years later. *Archives of Pediatrics and Adolescent Medicine 162*(12), 1133–1141.

Heckman, J. and R. Pinto (2013). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. Forthcoming, *Econometric Reviews*.

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

Heckman, J. J., J. E. Humphries, and T. Kautz (2014a). The economic and social benefits of GED certification. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press.

Heckman, J. J., J. E. Humphries, and T. Kautz (Eds.) (2014b). *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press.

Heckman, J. J., J. E. Humphries, S. Urzúa, and G. Veramendi (2011). The effects of educational choices on labor market, health, and social outcomes. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J. and T. Kautz (2012, August). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464. Adam Smith Lecture.

Heckman, J. J. and T. Kautz (2014a). Achievement tests and the role of character in American life. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of*

Achievement Tests: The GED and the Role of Character in American Life. Chicago: University of Chicago Press.

Heckman, J. J. and T. Kautz (2014b). Fostering and measuring skills: Interventions that improve character and cognition. In *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago, IL: University of Chicago Press. Forthcoming.

Heckman, J. J. and D. V. Masterov (2007). The productivity argument for investing in young children. *Review of Agricultural Economics 29*(3), 446–493.

Heckman, J. J., S. H. Moon, and R. Pinto (2012). The effects of early intervention on abilities and social outcomes: Evidences from the Carolina Abecedarian study. Unpublished manuscript, University of Chicago.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, August). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

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

Heckman, J. J. and S. Mosso (2013). The economics of human development and social mobility. Unpublished manuscript, University of Chicago, Department of Economics. Under review, *Annual Review of Economics*.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 1–35.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006, July). The effects of cognitive and noncog-

nitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heine, S. J., E. E. Buchtel, and A. Norenzayan (2008). What do cross-national comparisons of personality traits tell us? The case of Conscientiousness. *Psychological Science 19*(4), 309–313.

Heine, S. J., S. Kitayama, D. R. Lehman, T. Takata, E. Ide, C. Leung, and H. Matsumoto (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology 81*(4), 599–615.

Heine, S. J., D. R. Lehman, K. Peng, and J. Greenholtz (2002, June). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology 82*(6), 903–918.

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

Hieronymus, A. N., E. F. Lindquist, and H. D. Hoover (1980). *Iowa Tests of Basic Skills: Primary Battery*. Iowa City, IA: Houghton.

Hillage, J., J. Regan, J. Dickson, and K. McLoughlin (2002). Employers skill survey: 2002. Research Report RR372, Department for Education and Skills.

Hirschi, T. and M. Gottfredson (1993). Commentary: Testing the general theory of crime. *Journal of Research in Crime and Delinquency 30*(1), 47–54.

Hogan, J. and B. Holland (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology 88*(1), 100–112.

Hollenbeck, K. (2008). State use of workforce system net impact estimates and rates of

return. Technical report, Upjohn Institute. Presented at the Association for Public Policy Analysis and Management (APPAM) Conference, Los Angeles, CA.

Holmlund, H. and O. Silva (2009). Targeting non-cognitive skills to improve cognitive outcomes: Evidence from a remedial education intervention. Discussion Paper 4476, IZA.

Holt, M. M. and T. R. Hobbs (1979). The effects of token reinforcement, feedback and response cost on standardized test performance. *Behaviour Research and Therapy 17*(1), 81–83.

Holzer, H. (1997). Is there a gap between employer skill needs and the skills of the work force? In A. Lesgold, M. Feuer, and A. Black (Eds.), *Transitions in Work and Learning: Implications for Assessment*, Chapter 2, pp. 6–33. Washington, DC: National Academy Press.

Howard, K. S. and J. Brooks-Gunn (2009). The role of home-visiting programs in preventing child abuse and neglect. *The Future of Children 19*(2), 119–146.

Huang, D., B. Gribbons, K. S. Kim, C. Lee, and E. L. Baker (2000). A decade of results: The impact of the LA's BEST after school enrichment program on subsequent student achievement and performance. Technical report, UCLA Center for the Study of Evaluation, Graduate School of Education and Information Studies, Los Angeles, CA.

Huang, D., K. S. Kim, A. Marshall, and P. Pérez (2005). Keeping kids in school: An LA's BEST example—a study examining the long-term impact of LA's BEST on students' dropout rates. Technical report, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles, CA.

Hunt, E. (2012). Improving intelligence. Presented at the American Psychological Association Annual Meeting, Sunday, May 27, 2012, Chicago, IL.

Jablonka, E. and G. Raz (2009). Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology 84*(2), pp. 131–176.

Jackson, C. K. (2013). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. Working Paper 18624, NBER.

Jencks, C. (1979). *Who Gets Ahead? The Determinants of Economic Success in America*. New York: Basic Books.

John, O. P., A. Caspi, R. W. Robins, and T. E. Moffitt (1994). The "little five": Exploring the nomological network of the five-factor model of personality in adolescent boys. *Child Development 65*(1), 160–178.

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

Johnson, A. W. (1999). Assessing the impact of the Sponsor-A-Scholar youth mentoring program on student performance. Technical Report 355, Mathematica Policy Research.

Johnson, D. and T. Walker (1991). A follow-up evaluation of the Houston Parent-Child Development Center: School performance. *Journal of Early Intervention 15*(3), 226–236.

Jolly, R. (2007). Early childhood development: The global challenge. *The Lancet 369*(9555), 8–9.

Kaestle, C. F. (1984, May). Moral education and common schools in America: A historian's view. *Journal of Moral Education 13*(2), 101–111.

Kagitcibasi, C., D. Sunar, and S. Bekman (2001). Long-term effects of early intervention: Turkish low-income mothers and children. *Journal of Applied Developmental Psychology 22*(4), 333–361.

Kagitcibasi, C., D. Sunar, S. Bekman, N. Baydar, and Z. Cemalcilar (2009). Continuing effects of early enrichment in adult life: The Turkish Early Enrichment Project 22 years later. *Journal of Applied Developmental Psychology 30*(6), 764–779.

Kahne, J. and K. Bailey (1999). The role of social capital in youth development: The case of "I Have a Dream" programs. *Educational Evaluation and Policy Analysis 21*(3), 321–343.

Kemple, J. and J. C. Snipes (2000). Career Academies: Impacts on student engagement and performance in high school. Report 03/2000, MDRC.

Kemple, J. and C. Willner (2008). Career Academies: long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood. Report 06/2008, MDRC.

Kern, M. L. and H. S. Friedman (2008). Do conscientious individuals live longer? *Health Psychology 27*(5), 505–512.

Kettner, N. (1976). Armed Services Vocational Aptitude Battery (ASVAB form 5): Comparison with GATB and DAT tests: Final report for period May 1975–October 1976. Technical report, DTIC Document: Department of Defense, Department of the Air Force, Air Force Systems Command, Air Force Human Resources Laboratory.

Kitzman, H., D. L. Olds, R. Cole, C. Hanks, E. Anson, K. Arcoleo, D. W. Luckey, M. Knudtson, C. R. Henderson, and J. Holmberg (2010, May). Enduring effects of prenatal and infancy home visiting by nurses on children: Follow-up of a randomized trial among children at age 12 years. *Journal of the American Medical Association 164*(5), 412–418.

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.

Krueger, A. B. (2003, February). Economic considerations and class size. *Economic Journal 113*(485), F34–F63.

Kubzansky, L. D., D. Sparrow, P. Vokonas, and I. Kawachi (2001). Is the glass half empty or half full? A prospective study of optimism and coronary heart disease in the normative aging study. *Psychosomatic Medicine 63*(6), 910–916.

Lake, A. (2011). Early childhood development—global action is overdue. *The Lancet 378*(9799), 1277–1278.

Lally, J. R., P. L. Mangione, and A. S. Honig (1987). *The Syracuse University Family Development Research Program: Long-Range Impact of an Early Intervention with Low-Income Children & their Families*. San Francisco, CA: Center for Child and Family Studies, Far West Laboratory for Educational Research and Development.

Lang, F. R., D. John, O. Lüdtke, J. Schupp, and G. G. Wagner (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior research methods 43*(2), 548–567.

Lapsley, D. and D. Yeager (2012). Moral-character education. In I. Weiner, W. Reynolds, and G. Miller (Eds.), *Handbook of Psychology*, Volume 7 of *Educational Psychology*, Chapter 7, pp. 117–146. New York, NY: Wiley.

Larson, G. E., D. P. Saccuzzo, and J. Brown (1994). Motivation: Cause or confound in information processing/intelligence correlations? *Acta Psychologica 85*(1), 25–37.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. New York: Farrar, Straus and Giroux.

Lerman, R. I. (2013). Are employability skills learned in U.S. youth education and training programs? Under review, *IZA Journal of Labor Policy*.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology 22*(140), 55.

Lillard, A. and N. Else-Quest (2006). The early years: Evaluating Montessori. *Science 313*(5795), 1893–1894.

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research 37*(3), 888–902.

Ludwig, J. and D. L. Miller (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity approach. *Quarterly Journal of Economics 122*(1), 159–208.

Lutz, M. N., J. Fantuzzo, and P. McDermott (2000). *Adjustment Scales for Preschool Intervention*. Philadelphia, PA: University of Pennysylvania.

Mallar, C., S. Kerachsky, C. Thornton, and D. Long (1982). Evaluation of the impact of the Job Corps program: Third follow-up report. Technical Report PR82-05, Mathematica Policy Research, Inc., Princeton, NJ.

Martin, L. R., H. S. Friedman, and J. E. Schwartz (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology 26*(4), 428–436.

Martins, P. S. (2010). Can targeted, non-cognitive skills programs improve achievement? Discussion Paper 5266, IZA.

McAdams, D. P. (2006). *The Person: A New Introduction to Personality Psychology* (4 ed.). Hoboken, N.J.: J. Wiley and Sons.

McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist 33*(3), 284–289.

McCord, J. and W. McCord (1959). A follow-up report on the Cambridge-Somerville Youth Study. *Annals of the American Academy of Political and Social Science 322*(1), 89–96.

McCormick, M. C., J. Brooks-Gunn, S. L. Buka, J. Goldman, J. Yu, M. Salganik, D. T. Scott, F. C. Bennett, L. L. Kay, J. C. Bernbaum, C. R. Bauer, C. Martin, E. R. Woods, A. Martin, and P. H. Casey (2006, March). Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics 117*(3), 771–780.

McDowell, J., D. Oliver, M. Persson, R. Fairbrother, S. Wetzlar, J. Buchanan, and T. Shipstone (2011). A shared responsibility: Apprenticeships for the 21st century. `http://www.australianapprenticeships.gov.au/faq/documents/apprenticeshipsforthe21stcenturyexpertpanel.pdf`, last accessed February 11, 2013.

McLanahan, S. (2004, November). Diverging destinies: How children are faring under the second demographic transition. *Demography 41*(4), 607–627.

Michalopoulos, C., D. Tattrie, C. Miller, P. K. Robins, P. Morris, D. Gyarmati, C. Redcross, K. Foley, and R. Ford (2002). Final report on the Self-Sufficiency Project for long-term welfare recipients. Report, Social Research and Demonstration Corporation.

Millenky, M., D. Bloom, and C. Dillon (2010). Making the transition: Interim results of the National Guard Youth ChalleNGe evaluation. Report 05/2010, MDRC.

Millenky, M., D. Bloom, S. Muller-Ravett, and J. Broadus (2011). Staying on course: Three-year results of the National Guard Youth ChalleNGe evaluation. Report 06/2011, MDRC.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Mischel, W., O. Ayduk, M. G. Berman, B. J. Casey, I. H. Gotlib, J. Jonides, E. Kross, T. Teslovich, N. L. Wilson, V. Zayas, and Y. Shoda (2011). 'Willpower' over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience 6*(2), 252–256.

Moon, S. H. (2012, July). Time to invest in disadvanted young children. *Samsung Economic Research Institute Quarterly 5*(3), 50–59.

Mroczek, D. K. and A. Spiro (2007). Personality change influences mortality in older men. *Psychological Science 18*(5), 371–376.

Murray, H. A. (1938). *Explorations in Personality: A Clinical and Experimental Study of Fifty Men of College Age*. New York: Oxford University Press.

Myers, R. (1992). *The Twelve Who Survive: Strengthening Programmes of Early Childhood Development in the Third World*. New York, NY: Routledge/UNESCO.

National Apprenticeship Service (2012). History of apprenticeships. `http://www.apprenticeships.org.uk/About-Us/History-of-Apprenticeships.aspx`, last accessed February 11, 2013.

National Centre for Vocational Education Research (2012). Historical time series of apprenticeships and traineeships in Australia from 1963. http://www.ncver.edu.au/publications/2329.html.

Naumann, L. P. and O. P. John (2011). Toward a domain-specific approach to cultural differences: The influence of cultural values and reference-group standards on self-reported personality. Unpublished manuscript, Sonoma State University, Department of Psychology.

Neal, D. A. (1997, January). The effects of Catholic secondary schooling on educational achievement. *Journal of Labor Economics 15*(1, Part 1), 98–123.

Nelsen, B. (1997). Should social skills be in the vocational curriculum? Evidence from the automotive career field. In A. M. Lesgold, M. J. Feuer, and A. M. Black (Eds.), *Transitions in Work and Learning: Implications for Assessment, Papers and Proceedings*. Washington, DC: National Academy Press.

Niles, M. D., A. J. Reynolds, and M. Nagasawa (2006). Does early childhood intervention

affect the social and emotional development of participants? *Early Childhood Research and Practice 8*(1).

Nisbett, R. E. (2009, February). *Intelligence and How to Get It: Why Schools and Cultures Count*. New York, NY: W. W. Norton and Company.

Nisbett, R. E., J. Aronson, C. Blair, W. Dickens, J. Flynn, D. F. Halpern, and E. Turkheimer (2012). Intelligence: New findings and theoretical developments. *American Psychologist 67*(2), 130–159.

Nyhus, E. K. and E. Pons (2005). The effects of personality on earnings. *Journal of Economic Psychology 26*(3), 363–384.

Olds, D. L. (2006). The Nurse-Family Partnership: An evidence-based preventive intervention. *Infant Mental Health Journal 27*(1), 5–25.

Olds, D. L., H. Kitzman, R. Cole, C. Hanks, K. Arcoleo, E. Anson, D. W. Luckey, M. Knudtson, C. R. Henderson, J. Bondy, and A. J. Stevenson (2010, May). Enduring effects of prenatal and infancy home visiting by nurses on maternal life course and government spending: Follow-up of a randomized trial among children at age 12 years. *Journal of the American Medical Association 164*(5), 419–424.

Olds, D. L., H. Kitzman, R. Cole, J. Robinson, K. Sidora, D. W. Luckey, C. R. Henderson, C. Hanks, J. Bondy, and J. Holmberg (2004, December). Effects of nurse home-visiting on maternal life course and child development: Age 6 follow-up results of a randomized trial. *Pediatrics 114*(6), 1550–1559.

Olds, D. L., H. Kitzman, C. Hanks, R. Cole, E. Anson, K. Sidora-Arcoleo, D. W. Luckey, C. R. Henderson, J. Holmberg, R. A. Tutt, A. J. Stevenson, and J. Bondy (2007, December). Effects of nurse home-visiting on maternal and child functioning: Age 9 follow-up of a randomized trial. *Pediatrics 120*(4), 832–845.

Olds, D. L., J. Robinson, L. Pettitt, D. W. Luckey, J. Holmberg, R. K. Ng, K. Isacks, K. Sheff, and C. R. Henderson (2004, December). Effects of home visits by paraprofessionals and by nurses: Age 4 follow-up results of a randomized trial. *Pediatrics 114*(6), 1560–1568.

Organisation of Economic Cooperation and Development (2001). OECD employment and labour market statistics. Data available from `http://www.oecd-ilibrary.org/employment/data/oecd-employment-and-labour-market-statistics_lfs-data-en`.

Orr, L., H. Bloom, S. Bell, W. Lin, G. Cave, and F. Doolittle (1994). *The National JTPA Study. Impacts, Benefits, and Costs of Title II-A*. Bethesda, MD: Abt Associates.

Peng, K., R. E. Nisbett, and N. Y. Wong (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods 2*(4), 329–344.

Petersen, A. C., J. E. Schulenberg, R. H. Abramowitz, D. Offer, and H. D. Jarcho (1984). A Self-Image Questionnaire for Young Adolescents (SIQYA): Reliability and validity studies. *Journal of Youth and Adolescence 13*(2), 93–111.

Pratt, T. C. and F. T. Cullen (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology 38*(3), 931–964.

Quinn, L. M. (2014). An institutional history of the GED. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press.

Rauner, F. (2007). Vocational education and training—A European perspective. In A. Brown, S. Kirpal, and F. Rauner (Eds.), *Identities at Work*, Volume 5 of *Technical and Vocational Education and Training: Issues, Concerns and Prospects*, pp. 115–144. Springer.

Raven, J., J. C. Raven, and J. Court (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Raven, J. C. (1962). *Advanced Progressive Matrices: Sets I and II* (Revised ed.). London: H.K. Lewis.

Reed, D., A. Yung-Hsu Liu, R. Kleinman, A. Mastri, D. Reed, S. Sattar, and J. Ziegler (2012). An effectiveness assessment and cost-benefit analysis of registered apprenticeship in 10 states. Mathematica Final Report 06689.090 and 40096, Mathematica Policy Research.

Reynolds, A. J. (1994). Effects of a preschool plus follow-on intervention for children at risk. *Developmental Psychology 30*(6), 787–804.

Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly 10*(1), 1–31.

Reynolds, A. J. (2000). *Success in Early Intervention: The Chicago Child-Parent Centers*. Lincoln: University of Nebraska Press.

Reynolds, A. J. and J. A. Temple (1998, February). Extended early childhood intervention and school achievement: Age 13 findings from the Chicago Longitudinal Study. *Child Development 69*(1), 231–246.

Reynolds, A. J., J. A. Temple, S.-R. Ou, I. A. Arteaga, and B. A. B. White (2011, July). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science 333*(6040), 360–364.

Reynolds, A. J., J. A. Temple, D. L. Robertson, and E. A. Mann (2001, May). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest—A 15-year follow-up of low-income children in public schools. *Journal of American Medical Association 285*(18), 2339–2346.

Reynolds, A. J., J. A. Temple, D. L. Robertson, and E. A. Mann (2002). Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis 24*(4), 267–303.

Reynolds, A. J., J. A. Temple, B. A. White, S.-R. Ou, and D. L. Robertson (2011, January–February). Age 26 cost-benefit analysis of the Child-Parent Center early education program. *Child Development 82*(1), 379–404.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality 43*(2), 137–145.

Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg (2007, December). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science 2*(4), 313–345.

Roberts, R. D., G. N. Goff, F. Anjoul, P. C. Kyllonen, G. Pallier, and L. Stankov (2000). The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and Individual Differences 12*(1), 81–103.

Roder, A. and M. Elliot (2011). A promising start: Year-Up's initial impacts on young adults' careers. Technical report, Economic Mobility Corporation.

Rodríguez-Planas, N. (2010). Mentoring, educational services, and economic incentives: Longer-term evidence on risky behaviors from a randomized trial. Discussion Paper 4968, IZA.

Rodríguez-Planas, N. (2012, October). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. *American Economic Journal: Applied Economics 4*(4), 121–139.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Rothstein, R., R. Jacobsen, and T. Wilder (2008). *Grading Education: Getting Accountability Right*. New York: Economic Policy Institute and Teachers College Press.

Rutter, M. (2006). Implications of resilience concepts for scientific understanding. *Annals of the New York Academy of Sciences 1094*(1), 1–12.

Ryan, R. M. and E. L. Deci (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology 25*(1), 54–67.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology 82*(1), 30–43.

Schmidt, F. L. and J. Hunter (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology 86*(1), 162–173.

Schmitt, D. P. (2002). Are sexual promiscuity and relationship infidelity linked to different personality traits across cultures? Findings from the International Sexuality Description Project. *Online Readings in Psychology and Culture 4*(Unit 4), 1–22. Retrieved from http://scholarworks.gvsu.edu/orpc/vol4/iss4/4.

Schmitt, D. P., J. Allik, R. R. McCrae, and V. Benet-Martínez (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology 38*(2), 173–212.

Schochet, P. Z., J. Burghardt, and S. Glazerman (2001). *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*. Princeton, NJ: Mathematica Policy Research, Inc.

Schochet, P. Z., J. Burghardt, and S. McConnell (2008, December). Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review 98*(5), 1864–1886.

Schulz, R., J. Bookwala, J. E. Knapp, M. Scheier, and G. M. Williamson (1996). Pessimism, age, and cancer mortality. *Psychology and Aging 11*(2), 304–309.

Schumacher, R. (2003). Family support and parent involvement in Head Start: What do Head Start program performance standards require? Technical report, Center for Law and Social Policy.

Scribner, J. P. and D. Wakelyn (1998). Youth apprenticeship experiences in Wisconsin: A stakeholder-based evaluation. *The High School Journal 82*(1), 24–34.

Secretary's Commission on Achieving Necessary Skills (1992). *Learning a Living: A Blueprint for High Performance: A SCANS Report for America 2000*. Washington, DC: United States Government Printing Office.

Segal, C. (2008, October). Working when no one is watching: Motivation, test scores, and economic success. Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain.

Shure, M. B. and G. Spivack (1988). Interpersonal cognitive problem solving. In R. H. Price, E. L. Cowen, R. P. Lorion, and J. Ramos-McKay (Eds.), *14 Ounces of Prevention: A Casebook for Practitioners*, Chapter 6, pp. 69–82. Washington, DC: American Psychological Association.

Social and Character Development Research Consortium (2010, October). Efficacy of school-wide programs to promote social and character development and reduce problem behavior in elementary school children. Research Report NCER 20112001, National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Sroufe, L. A. (1997, Spring). Psychopathology as an outcome of development. *Development and Psychopathology 9*(2), 251–268.

Sroufe, L. A., B. Egeland, E. Carlson, and W. A. Collins (2005). *The Development of the Person: The Minnesota Study of Risk and Adaptation from Birth to Adulthood*. New York, NY: Guilford Press.

Stasz, C. (2001). Assessing skills for work: Two perspectives. *Oxford Economic Papers 53*(3), 385–405.

Sylva, K. (1997). The quest for quality in curriculum. In L. J. Schweinhart and D. P. Weikart (Eds.), *Lasting Differences: The High/Scope Preschool Curriculum Comparison Study through Age 23*, pp. 89–93. Ypsilanti: High/Scope Press.

Temple, J. A. and A. J. Reynolds (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and related programs. *Economics of Education Review 26*(1), 126–144.

Terman, L. M., B. T. Baldwin, E. Bronson, J. C. DeVoss, F. Fuller, T. Lee Kelley, M. Lima, H. Marshall, A. H. Moore, A. S. Raubenheimer, G. M. Ruch, R. L. Willoughby, J. Benson Wyman, and D. Hazeltine Yates (1925). *Genetic Studies of Genius: Mental and Physical Traits of a Thousand Gifted Children*, Volume 1. Stanford University, CA: Stanford University Press.

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

Terman, L. M., M. H. Oden, N. Bayley, H. Marshall, Q. McNemar, and E. B. Sullivan (1947). *Genetic Studies of Genius: The Gifted Child Grows Up: Twenty-Five Years' Follow-Up of a Superior Group*, Volume 4. Stanford University, CA: Stanford University Press.

Thaler, R. H. (2008, July). A short course in behavioral economics. Edge Master Class, Sonoma, CA, July 25–27, 2008.

The KIPP Foundation (2011). The promise of college completion: KIPP's early successes and challenges. Report, The KIPP Foundation, San Francisco, CA.

Tierney, J. P., J. Baldwin-Grossman, and N. L. Resch (1995). Making a difference: An impact study of Big Brothers Big Sisters. Report, Public/Private Ventures.

Tough, P. (2012). *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. Boston, MA: Houghton Mifflin Harcourt.

Tuttle, C. C., B.-r. Teh, I. Nichols-Barrer, B. P. Gill, and P. Gleason (2010). Student characteristics and achievement in 22 KIPP middle schools. Report 06441.900, Mathematica.

Tyler, R. W. (1973). Assessing educational achievement in the affective domain. *Measurement in Education 4*(3), 1–8.

Walker, G. C. and F. Vilella-Velez (1992). *Anatomy of a Demonstration: The Summer Training and Education Program (STEP) from Pilot through Replication and Postprogram Impacts*. Philadelphia, PA: Public/Private Ventures.

Walker, S., T. Wachs, S. Grantham-McGregor, M. Black, C. Nelson, S. Huffman, H. Baker-Henningham, S. Chang, J. Hamadani, B. Lozoff, J. Meeks Gardner, C. Powell, A. Rahman, and L. Richter (2011). Inequality in early childhood: risk and protective factors for early child development. *The Lancet 378*(9799), 1325–1338.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet 366*(9499), 1804–1807.

Walker, S. P., S. M. Chang, C. A. Powell, E. Simonoff, and S. M. Grantham-McGregor (2007). Early childhood stunting is associated with poor psychological functioning in late adolescence and effects are reduced by psychosocial stimulation. *The Journal of Nutrition 137*(11), 2464–2469.

Walker, S. P., S. M. Chang, M. Vera-Hernández, and S. Grantham-McGregor (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics 127*(5), 849–857.

Walker, S. P., T. D. Wachs, J. M. Gardner, B. Lozoff, G. A. Wasserman, E. Pollitt, J. A. Carter, and The International Child Development Steering Group (2007, January). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet 369*(9556), 145–157.

Walsh, D. A. (2005). *Why Do They Act That Way? A Survival Guide to the Adolescent Brain for You and Your Teen* (1 ed.). New York: Free Press.

Washington Workforce Training Board (2008). Washington state employers workforce needs and practices survey. Statewide report.

Webb, E. (1915). Character and intelligence. *British Journal of Psychology Supplement 1*(3), 1–99.

Westat (2010). Head Start Impact Study: Final report. Contract 282-00-0022, U.S. Department of Health and Human Services.

Westwood, A. (2004). Skills that matter and shortages that don't. In C. Warchust, I. Grugulis, and E. Keep (Eds.), *The Skills that Matter*. New York, NY: Palgrave-Macmillan.

Winkelmann, R. (1996, July). Employment prospects and skill acquisition of apprenticeship-trained workers in Germany. *Industrial and Labor Relations Review 49*(4), 658–672.

Wolter, S. C. and P. Ryan (2011). Apprenticeship. In S. M. Eric A. Hanushek and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3 of *Handbooks in Economics*, Chapter 11, pp. 521–576. Elsevier.

Young, M. E. (1996). *Early Childhood Development: Investing in the Future*. Washington, DC: World Bank.

Zabel, J., S. Schwartz, and S. Donald (2010). The impact of the Self-Sufficiency Project

on the employment behaviour of former welfare recipients. *Canadian Journal of Economics 43*(3), 882–918.

Zabel, J., S. Schwartz, and S. Donald (2013, February). An analysis of the impact of the Self-Sufficiency Project on wages. *Empirical Economics 44*(1), 231–259.

Zemsky, R. (1997). Skills and the economy: An employer context for understanding the school-to-work transition. In A. Lesgold, M. Feuer, and A. Black (Eds.), *Transitions in Work and Learning: Implications for Assessment*. Washington, DC: National Academy Press.