Heckman, James J.; Kautz, Tim

Working Paper
Hard evidence on soft skills

Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, No. 6580

Provided in Cooperation with:
Institute of Labor Economics (IZA)


This Version is available at:
http://hdl.handle.net/10419/58704

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Hard Evidence on Soft Skills

James J. Heckman
Tim Kautz

May 2012
Hard Evidence on Soft Skills

James J. Heckman
University of Chicago, American Bar Foundation,
University College Dublin and IZA

Tim Kautz
University of Chicago

Discussion Paper No. 6580
May 2012

IZA
P.O. Box 7240
53072 Bonn
Germany
Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.
ABSTRACT

Hard Evidence on Soft Skills*

This paper summarizes recent evidence on what achievement tests measure; how achievement tests relate to other measures of “cognitive ability” like IQ and grades; the important skills that achievement tests miss or mismeasure, and how much these skills matter in life. Achievement tests miss, or perhaps more accurately, do not adequately capture, soft skills – personality traits, goals, motivations, and preferences that are valued in the labor market, in school, and in many other domains. The larger message of this paper is that soft skills predict success in life, that they causally produce that success, and that programs that enhance soft skills have an important place in an effective portfolio of public policies.

JEL Classification: I20, D01

Keywords: personality, achievement tests, IQ, cognition

Corresponding author:

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
USA
E-mail: jjh@uchicago.edu

* This paper was presented as the Adam Smith Lecture at the Annual Meeting of the European Association of Labour Economists held in Cyprus, September 2011. This research was supported in part by the University of Chicago, A New Science of Virtues: A Project of the University of Chicago, the American Bar Foundation, a conference series from the Spencer Foundation, the JB & MK Pritzker Family Foundation, Susan Thompson Buffett Foundation, the Geary Institute, University College Dublin, Ireland, NICHD R37 HD065072 and R01 HD054702, we acknowledge the support of a European Research Council grant hosted by University College Dublin, DEVHEALTH 269874, a grant to the Becker Friedman Institute for Research and Economics from the Institute for New Economic Thinking (INET), and an anonymous funder. 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 for research assistance. Paul Devereux made helpful comments that are incorporated in this paper.
1 Introduction

Contemporary society places great value on standardized achievement tests to sift and sort people, to evaluate schools, and to assess the performance of nations. Admissions committees use tests like the SAT (Scholastic Assessment Test), the ACT (American College Testing), and the GRE (Graduate Record Examinations) to screen applicants. In the United States, the No Child Left Behind (NCLB) Act stipulates that government-run schools must administer standardized achievement tests in order to be eligible for federal funding. The Programme for International Student Assessment (PISA) evaluates student performance in math, science, and reading across countries. The results attract media attention and influence policy. The year 2000 PISA test results caused Germany to re-evaluate its educational system and introduce a variety of educational reforms (Grek, 2009).

Despite the widespread use of standardized achievement tests, the traits that they measure are not well-understood. This paper summarizes recent evidence on what achievement tests capture; how achievement tests relate to other measures of “cognitive ability” like IQ and grades; the important skills that achievement tests miss or mismeasure, and how much these other skills matter in life.

Achievement tests miss, or more accurately, do not adequately capture, soft skills — personality traits, goals, motivations, and preferences that are valued in the labor market, in school, and in many other domains. The larger message of this paper is that soft skills predict success in life, that they produce that success, and that programs that enhance soft skills have an important place in an effective portfolio of public policies.

Measurement of cognition and educational attainment has been refined during the past century. Psychometricians have shown that cognitive ability has multiple facets. This

---

1 Sales of achievement tests have increased by nearly 400% between 1959 and 2005 (Digest of Education Statistics, various years; The Bowker Annual: Library and Book Trade Almanac, various years).

2 This paper draws on and supplements Borghans et al. (2008), Almlund et al. (2011), and Heckman et al. (2012).

3 See Carroll (1994) and Ackerman and Heggestad (1997) for a discussion.
progress is not widely appreciated. Many social scientists—even many psychologists—continue to use IQ tests, standardized achievement tests, and grades interchangeably to proxy “cognitive ability.”

Even though scores on IQ tests, standardized achievement tests, and grades are positively correlated with each other, the recent literature shows that they measure different skills and depend on different facets of cognitive ability. Recent research also shows that all three measures are associated with personality, but to different degrees across various cognitive measures.

Standardized achievement tests were designed to capture “general knowledge” produced in schools and through life experiences. Such knowledge is thought to be relevant to success inside and outside of the classroom. However, achievement tests are often validated using other standardized achievement tests or other measures of cognitive ability—surely a circular practice.

A more relevant validity criterion is how well these tests predict meaningful outcomes, such as educational attainment, labor market success, crime, and health. No single measure of cognitive ability predicts much of the variance in these outcomes, and measurement error does not account for most of the remaining variance, leaving much room for other determinants of success.

Success in life depends on personality traits that are not well captured by measures of cognition. Conscientiousness, perseverance, sociability, and curiosity matter. While economists have largely ignored these traits, personality psychologists have studied them over the last century. They have constructed measures of them and provide evidence that these traits predict meaningful life outcomes.

Many scholars—inside and outside of psychology—have questioned the existence of stable personality traits, arguing that constraints and incentives in situations almost entirely de-

---

4 Many call this “IQ”, e.g., Flynn (2007), Nisbett (2009), and Nisbett et al. (2012).
5 On the magnitudes of measurement error on a variety of economic measures, see Bound et al. (2001). These authors report that at most 15–30% of earnings variance is due to measurement error.
6 Some early studies in economics are Bowles and Gintis (1976), and Bowles et al. (2001). An important study in sociology is Jencks (1979). Work in psychology going back to Terman et al. (1925) shows that personality traits matter (see Murray 1938, Terman et al. 1947, and the discussion in Gensowski 2012).
terminate behavior. These scholars claim that people are like chameleons—they are what they are shaped to be in any situation. A substantial body of evidence shows that stable traits exist. People tend to behave in the same fashion across a wide range of situations. Evidence from genetics and neuroscience provides a biological basis for the existence of such traits, suggesting that something tied to the person, not the just the situation, affects behavior.

Throughout this paper we use the term “personality traits” to describe the personal attributes not thought to be captured by measures of abstract reasoning power. 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. The term “traits” suggests a sense of permanence and possibly also of heritability. The terms “skills” and “character” suggest that they can be learned. In reality, the extent to which these personal attributes can change lies on a spectrum. Both cognitive and personality traits can change and be changed over the life cycle but through different mechanisms and to different degrees at different ages. To avoid confusion, throughout this paper we use the term “trait” to capture the set of personal attributes we study.

Psychological traits are not directly observed. There is no ruler for perseverance, no caliper for intelligence. All cognitive and personality traits are measured using performance on “tasks,” broadly defined. Different tasks require different traits in different combinations. Some distinguish between measurements of traits and measurements of outcomes, but this distinction is misleading. Both traits and outcomes are measured using performance on some task or set of tasks.

Psychologists sometimes claim to circumvent this measurement issue by creating tax-
onomies of traits and by applying intuitive names to responses on questionnaires. These
questionnaires are not windows to the soul. They are still rooted in task performance or
behavior. Responding to a questionnaire is a task. Many of the questionnaires inquire di-
rectly about behavior, e.g. a measure of Agreeableness used in the German Socioeconomic
Panel asks the extent to which a respondent “is sometimes somewhat rude to others.” How
else can one answer that question but reflect on one’s behavior? IQ tests and standardized
achievement tests also measure performance on different “cognitively loaded” tasks.

Performance on most tasks depends on effort, personality traits, cognitive ability, and
incentives, although the importance of each differs by task. This dependence creates a
fundamental problem in measuring traits. Most studies in psychology devise a set of measures
to capture a trait but do not standardize for incentives in the situation in which the trait
is being measured or for other traits. Measured cognitive ability and measured personality
depend on a constellation of factors. The identification problem arising from the multiple
determinants of performance on tasks is empirically important, even for measures of cognitive
ability. Incentives can affect performance on IQ tests. Multiple traits affect performance on
cognitive tasks. For example, personality traits affect achievement test scores and grades.\footnote{See \cite{bo11a}.} Caution is required in taking the measures developed by psychologists too literally.

Nonetheless, measures of personality traits predict meaningful life outcomes. Conscien-
tiousness — the tendency to be organized, responsible, and hardworking—is the most widely
predictive of the commonly used personality measures. It predicts educational attainment,
health, and labor market outcomes as strongly as measures of cognitive ability.\footnote{See the evidence collected in \cite{al11,bo08,ro07}.}

Most studies in psychology only report correlations between measured traits and out-
comes without addressing whether the traits cause the outcomes and without controlling for
the other traits and incentives that determine performance on the tasks used to measure the
traits. While traits are relatively stable across situations, they are not set in stone. They
change over the life cycle. On average, Agreeableness and Conscientiousness tend to grow

\footnote{\cite{bo11a}.}
with age. Different facets of cognitive ability peak at different ages. Interventions, education, and parenting can affect traits in lasting ways.

This paper summarizes recent evidence that personality causally affects life outcomes. We review some of the literature from psychology and economics and then focus on two particularly compelling examples. First, we show how an achievement test, the General Educational Development (GED) test, fails to capture important traits that affect success in life. High school dropouts can take the GED to certify to employers and post-secondary institutions that their skills are equivalent to those of high school graduates who do not attend college. After accounting for differences in pre-existing cognitive ability, GED recipients perform much worse in the labor market than high school graduates and much more like other high school dropouts. GED recipients lack important personality traits. (See Heckman et al., 2011 and Heckman et al., 2012.)

Second, we show how an early childhood intervention, the Perry Preschool Program, improved the lives of disadvantaged children, even though the program did not permanently change the IQ of its participants. The program changed their personality traits in a lasting way. (See Heckman et al., 2012) Other interventions and observational studies provide supporting evidence that early-childhood investments improve outcomes through their effects on personality.

---

13 Borghans et al. (2008) and Amlund et al. (2011) present extensive surveys of this literature.

2 Defining and Measuring Personality Traits

2.1 History and Measurement of Cognitive Ability

Modern intelligence tests have been used for just over a century, beginning when a French minister of public instruction wished to identify retarded pupils in need of specialized education programs. In response, Alfred Binet created the first IQ test. IQ scores were interpreted as measuring a stable trait. The standardized achievement test was created in the wake of the perceived success of IQ tests as an objective and cost-effective measure of acquired skills. In contrast to IQ tests, standardized achievement tests were designed to measure “general knowledge” that could be acquired in schools and through life experiences and was widely applicable beyond the classroom to workplace and social functioning.

Achievement tests are typically validated on other achievement tests, IQ tests, and grades, rather than on tasks or outcomes in the labor market and in social functioning. Table 1 shows correlations among scores on standardized achievement tests, IQ tests, and grades. Standardized achievement tests are correlated with IQ tests, but the correlation depends on the subject area of the standardized achievement test. Hartlage and Steele (1977) find that the arithmetic portions of standardized achievement tests are the most highly correlated with IQ. Grades and scores on IQ tests and standardized achievement tests are far from perfectly correlated, suggesting that they measure different aspects of “cognitive functioning.”

Psychologists distinguish between fluid intelligence (the rate at which people learn) and crystalized intelligence (acquired knowledge). Achievement tests are heavily weighted to-

---

15 In 1904, La Société Libre pour l’Étude Psychologique de l’Enfant appointed a commission to create a mechanism for identifying these pupils in need of alternative education led by Binet. See Herrnstein and Murray (1994) for an overview of Binet’s life and work.

16 See Lindquist (1951).

17 It is an irony of the testing literature that high school grades are more predictive of first year college performance than SAT scores (Bowen et al., 2009). The SAT and related tests are thought to be more objective measures of student quality than high school grades (Lemann, 1999).

18 See e.g. Nisbett et al. (2012).
Table 1: Cognitive Ability Validities

<table>
<thead>
<tr>
<th>Test</th>
<th>Validation Domain</th>
<th>Estimate(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT (Achievement)</td>
<td>1st Year College GPA</td>
<td>0.35 - 0.53</td>
<td>Kobrin et al. (2008)</td>
</tr>
<tr>
<td>ACT (Achievement)</td>
<td>Early College GPA</td>
<td>0.42</td>
<td>ACT, Inc. (2007)</td>
</tr>
<tr>
<td>GED (Achievement)</td>
<td>HS Senior GPA</td>
<td>0.33 - 0.49</td>
<td>GED Testing Service (2009)</td>
</tr>
<tr>
<td>DAT (Achievement)</td>
<td>College GPA</td>
<td>0.13 - 0.62†</td>
<td>Omizo (1980)</td>
</tr>
<tr>
<td>AFQT (Achievement)</td>
<td>9th Grade GPA</td>
<td>0.54</td>
<td>Borghans et al. (2011a)</td>
</tr>
<tr>
<td>WAIS (IQ)</td>
<td>College GPA</td>
<td>0.38 - 0.43</td>
<td>Feingold (1982)</td>
</tr>
<tr>
<td>WAIS (IQ)</td>
<td>HS GPA</td>
<td>0.62</td>
<td>Feingold (1982)</td>
</tr>
<tr>
<td>Various IQ**</td>
<td>9th Grade GPA</td>
<td>0.42</td>
<td>Borghans et al. (2011a)</td>
</tr>
<tr>
<td>WISC (IQ)</td>
<td>WRAT (Achievement)</td>
<td>0.44 - 0.75†</td>
<td>Hartlage and Steele (1977)</td>
</tr>
<tr>
<td>WISC-R (IQ)</td>
<td>WRAT (Achievement)</td>
<td>0.35 - 0.76†</td>
<td>Hartlage and Steele (1977)</td>
</tr>
<tr>
<td>Various IQ**</td>
<td>AFQT (Achievement)</td>
<td>0.65</td>
<td>Borghans et al. (2011a)</td>
</tr>
<tr>
<td>Stanford Binet (IQ)</td>
<td>WISC-R (IQ)</td>
<td>0.77 - 0.87</td>
<td>Rothlisberg (1987), Greene et al. (1990)</td>
</tr>
<tr>
<td>Raven’s (IQ)</td>
<td>WAIS-R (IQ)</td>
<td>0.74 - 0.84</td>
<td>O’Leary et al. (1991)</td>
</tr>
<tr>
<td>WIAT (Achievement)</td>
<td>CAT/2 (Achievement)</td>
<td>0.69 - 0.83*</td>
<td>Michalko and Saklofske (1996)</td>
</tr>
</tbody>
</table>

† Large range is due to varying validity of eight subtests of DAT
‡ Ranges are given because correlations vary by academic subject
* Ranges are given because correlations vary by grade level
** IQ test scores in the NLSY79 are pooled across several IQ tests using IQ percentiles
wards crystallized intelligence, whereas IQ tests like Raven’s progressive matrices are heavily weighted toward fluid intelligence. Many psychologists do not recognize the differences among these measures and interchangeably use IQ, achievement tests, and grades to measure “cognitive ability” or “intelligence,” and this practice is also widespread in economics.

2.2 Defining and Measuring Psychological Traits

Validating one measure of cognitive ability using other measures of cognitive ability is circular. More relevant is how well these measures predict important life outcomes. Table 2 shows the extent to which IQ, standardized achievement tests, and grades explain the variance of outcomes at age 35 in the National Longitudinal Survey of Youth, 1979 (NLSY79) data. The three groups of columns under each category show results for different sub-samples based on the availability of the different cognitive measures. For each category, the first column shows the explained variance using only the designated measure of cognitive ability. Achievement tests and grades are more predictive than IQ. But none of these measures explains much of the variation of any outcome, leaving considerable room for other determinants. As noted in the introduction, it is unlikely that measurement error accounts for all of the remaining variance.

Personality is one missing ingredient. The second columns in each category preview our later discussion of the explanatory power of personality. They show the variance explained by measures of personality. In many cases, the variance explained by personality measures rivals that explained by measures of cognitive ability. The relative importance of person-

---

19 Roberts et al. (2000).
20 Raven et al. (1988). The high correlation between intelligence and achievement tests is in part due to the fact that both require cognitive ability and knowledge. Common developmental factors may affect both of these traits. Fluid intelligence promotes the acquisition of crystallized intelligence.
21 Carroll (1993) and Ackerman and Heggestad (1997) discuss more disaggregated facets of cognitive ability. See Flynn (2007) and Nisbett et al. (2012). For examples in economics see Benjamin et al. (2006).
22 See Almend and et al. (2011) for a more complete array of outcomes in the format of Table 2.
23 They include measures of adolescent risky behavior, self-esteem and locus of control (the extent to which people feel they have control over their lives). For precise definitions of the measures used, see the notes to Table 2.
ality depends on the outcome. The third column for each sub-sample shows the variance explained when both the cognitive and personality measures are used as predictors. In many cases, including the measures of personality in a regression with cognitive measures explains additional variance. The correlations between the set of measures of personality and the measures of cognition are positive, but not especially strong (see the bottom row of each table). Each set of traits has an independent influence on the outcomes in the table.

Even though economists have largely ignored personality traits, the pioneers of the original IQ tests recognized their importance. Alfred Binet the creator of the first IQ test (the Stanford-Binet test), noted that:

“[Success in school] ...admits of other things than intelligence; to succeed in his studies, one must have qualities which depend on attention, will, and character; for example a certain docility, a regularity of habits, and especially continuity of effort. A child, even if intelligent, will learn little in class if he never listens, if he spends his time in playing tricks, in giggling, in playing truant.”

(Binet and Simon 1916, p. 254)

Since the middle of the 19th century, personality psychologists have studied these traits. One leading personality psychologist defines personality traits in the following way:

“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, p. 140)

Personality traits are manifested through thoughts, feelings, and behaviors, and therefore, must be inferred from measures of performance on “tasks,” broadly defined. Under this

25 Lewis Terman, who created the Stanford-Binet test, even collected data on personality traits of a high-ability sample. In this sample, Conscientiousness is highly predictive of health and earnings (Savelyev 2011; Gensowski 2012).
Table 2: Predictive Validities in Outcomes that Matter (Adjusted R-Squared)

<table>
<thead>
<tr>
<th></th>
<th>IQ Sample</th>
<th></th>
<th>AFQT Sample</th>
<th></th>
<th>GPA Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Personality</td>
<td>Both</td>
<td>AFQT</td>
<td>Personality</td>
<td>Both</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males Earnings at Age 35</td>
<td>0.07</td>
<td>0.05</td>
<td>0.09</td>
<td>0.17</td>
<td>0.07</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.03</td>
<td>0.08</td>
<td>0.13</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.12</td>
<td>0.08</td>
<td>0.16</td>
<td>0.19</td>
<td>0.10</td>
<td>0.22</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females Earnings at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.09</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.10</td>
<td>0.08</td>
<td>0.14</td>
<td>0.17</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj, R^2 Cog, Personality</td>
<td>0.10</td>
<td>0.15</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: National Longitudinal Survey of Youth 1979. Table Description: The table shows the adjusted R-squared from regressions of later-life outcomes on measures of personality and cognition. For each cognitive measure, the first column shows the explained variance using only the measures of cognitive ability, the second column shows the explained variance without using only the measure of personality (Personality), and the third column shows the explained variance from using both the measures of personality and cognition (Both). The last row shows the adjusted R-squared from a regression of each cognitive measure on the personality measures. Measures of Personality and Cognition: The measures of personality include minor illegal activity in 1979 (vandalism, shoplifting, petty theft, fraud and fencing), major illegal activity in 1979 (auto theft, breaking/entering private property, grand theft), participation in violent crime in 1979 (fighting, assault and aggravated assault), tried marijuana before age 15, daily smoking before age 15, regular drinking before age 15, and any crime before age 15. It also includes measures of Self-Esteem and Locus of Control. Self-Esteem is measured using the ten-item Rosenberg scale administered in 1980. Locus of control is a measure of how much control an individual believes they have over their life and is measured using the 4-item Rotter scale. IQ and grades are from high school transcripts. IQ is pooled across several IQ tests using IQ percentiles. GPA is the individual's core-subject GPA from 9th grade. Outcomes: Due to the biennial nature of the survey after 1994, some respondents are not interviewed at age 35, for these individuals age 36 is used. Earnings includes zero-earners and excludes observations over $200,000 (2005 dollars). Hourly wage excludes observations less than $3 or over $200 (2005 dollars). Hours worked excludes observations less than 80 or more than 4000. Jail by age 35 indicates whether the respondent had listed residing in a jail or prison at some point before age 35. Welfare at age 35 indicates whether the respondent received any positive amount of welfare at age 35. Married at age 35 indicates whether the respondent was currently married. BA degree by age 35 indicates whether the respondent received a BA degree (or higher) by age 35. Depression in 1992 is based on the 7-item Center for Epidemiologic Studies Depression Scale (CES-D). Sample: The sample excludes the military over sample. The samples differ across the IQ, AFQT, and GPA due to missing measures across the samples.
definition, performance on IQ tests is a personality trait because it is an enduring pattern of behavior (how one “behaves” or “performs” on an IQ test).²⁶

Personality psychologists primarily measure personality traits using self-reported surveys. They have arrived at a relatively well-accepted taxonomy of traits called the “Big Five,” which include Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Table 3 defines these traits and their multiple facets.²⁷ Some argue that the Big Five are the longitude and latitude of personality, by which all more narrowly defined traits may be categorized (see Costa and McCrae 1992a). While the Big Five measures are now the most widely used, 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.²⁸ Other taxonomies, including psychopathology as measured by the DSM IV and measures of temperament, have also been related to the Big Five.²⁹ Almlund et al. (2011) show that economic preference parameters are not all that closely related to psychological traits and apparently represent different traits that, along with the psychological traits, govern behavior.

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 manifestation of a deeper set of preferences or goals. Achieving certain goals requires certain traits, e.g., a surgeon has to be conscientious and intelligent; a salesman has to be outgoing and engaging and so forth, etc. Under this view, traits are developed through practice, investment, and habituation. The deeper traits may be the preference parameters that generate the manifest traits. The apparent stability of expressed traits across situations may be a consequence of the stability of the goals and incentives to achieve these goals.³⁰

²⁶Studies of test-retest reliability of IQ tests show that scores are highly correlated across repeated testing occasions (see, e.g., Niolon 2005).
²⁷See e.g. Borghans et al. (2008).
²⁸See Borghans et al. (2008) and Almlund et al. (2011) for a comparison of these taxonomies.
²⁹See e.g. Cloninger et al. (1999).
³⁰McAdams (2006) adds goals to the list of possible traits. Almlund et al. (2011) develop a model in which preferences and traits determine the effort applied to tasks.
<table>
<thead>
<tr>
<th>Big Five Personality Factor</th>
<th>American Psychology Association Dictionary description</th>
<th>Facets (and correlated trait adjective)</th>
<th>Related Traits</th>
<th>Analogous Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>“the tendency to be organized, responsible, and hardworking”</td>
<td>Competence (efficient), Order (organized), Dutifulness (not careless), Achievement striving (ambitious), Self-discipline (not lazy), and Deliberation (not impulsive)</td>
<td>Grit, Perseverance, Delay of gratification, Impulse control, Achievement striving, Ambition, and Work ethic</td>
<td>Attention/(lack of) distractibility, Effortful control, Impulse control/delay of gratification, Persistence, Activity*</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>“the tendency to be open to new aesthetic, cultural, or intellectual experiences”</td>
<td>Fantasy (imaginative), Aesthetic (artistic), Feelings (excitable), Actions (wide interests), Ideas (curious), and Values (unconventional)</td>
<td>Sensory sensitivity, Pleasure in low-intensity activities, Curiosity</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>“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”</td>
<td>Warmth (friendly), Gregariousness (sociable), Assertiveness (self-confident), Activity (energetic), Excitement seeking (adventurous), and Positive emotions (enthusiastic)</td>
<td>Surgency, Social dominance, Social vitality, Sensation seeking, Shyness*, Activity*, Positive emotionality, and Sociability/affiliation</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>“the tendency to act in a cooperative, unselfish manner”</td>
<td>Trust (forgiving), Straightforwardness (not demanding), Altruism (warm), Compliance (not stubborn), Modesty (not show-off), and Tender-mindedness (sympathetic)</td>
<td>Empathy, Perspective taking, Cooperation, and Competitiveness</td>
<td>Irritability*, Aggressiveness, and Willfulness</td>
</tr>
<tr>
<td>Neuroticism/Emotional Stability</td>
<td>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.”</td>
<td>Anxiety (worrying), Hostility (irritable), Depression (not contented), Self-consciousness (shy), Impulsiveness (moody), Vulnerability to stress (not self-confident)</td>
<td>Internal vs. External, Locus of control, Core self-evaluation, Self-esteem, Self-efficacy, Optimism, and Axis I psychopathologies (mental disorders) including depression and anxiety disorders</td>
<td>Fearfulness/behavioral inhibition, Shyness*, Irritability*, Frustration (Lack of) soothability, Sadness</td>
</tr>
</tbody>
</table>

Notes: Facets specified by the NEO-PI-R personality inventory (Costa and McCrae 1992b). Trait adjectives in parentheses from the Adjective Check List (Gough and Heilbrun 1983). “These temperament traits may be related to two Big Five factors. Source: Table adapted from John and Srivastava (1999).
2.3 Identification Problems in Measuring Traits

Measuring traits is difficult, because, as suggested by Roberts’ definition of personality, all psychological measurements are calibrated on measured behavior, and the behaviors used to measure one trait can be influenced by incentives and other traits. To infer traits from behaviors requires standardizing for all of the other contributing factors that produce the observed behavior. The inability to parse and localize behaviors that depend on a single trait or ability gives rise to a fundamental identification problem that is typically ignored in empirical research investigating how psychological traits affect outcomes.31

There are two primary issues. First, behavior depends on incentives created in situations. Different incentives elicit different amounts of effort on the tasks used to measure traits. Accurately measuring personality traits requires standardizing for the effort applied in any task. Second, behavior in one task can depend on multiple traits. Not standardizing for incentives and other traits can produce misleading estimates of any trait.

These identification problems are empirically important when measuring any given trait. For example, incentives partly determine scores on IQ tests, even though some have argued that performance on IQ tests reflects maximal effort.32 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 incentivizing students with M&M candies.33 The incentives in one test do not affect performance on future tests.

The recent literature shows that personality traits are associated with standardized achievement test scores, which many analysts use interchangeably with IQ scores.34 Figures 1 and 2 show how the variance in the scores on two achievement tests, the Armed

31 See Borghans et al. (2011a) and Almlund et al. (2011).
32 A leading psychometrician, Carroll (1993), does not accept the notion that IQ captures maximal effort.
33 See Ayllon and Kelly (1972); Borghans et al. (2008); Breunung and Zella (1978); Clingman and Fowler (1976); Edlund (1972); Holt and Hobbs (1979); Larson et al. (1994); Segal (2008). This evidence is summarized in Borghans et al. (2008) and Almlund et al. (2011).
34 See e.g. Nisbett (2009).
Forces Qualifying Test (AFQT) and the closely related Differential Aptitudes Test (DAT) are decomposed into IQ and personality measures. Personality traits explain a substantial portion of the variances in both AFQT scores and DAT scores. The personality traits are incrementally valid in that they explain the variance above and beyond the variance that IQ explains when all three are included in a regression. These findings caution the interpretation that standardized achievement tests only measure cognitive ability. They are bundled with personality traits. In data from the Stella Maris secondary school in Maastricht, Holland, Openness to Experience is strongly correlated with IQ.

Figure 1: Decomposing Achievement Tests and Grades into IQ and Personality [NLSY79]

Further complicating identification, not everyone responds to incentives in the same way.

35 The correlation between DAT and AFQT scores in the National Longitudinal Study of Youth 1979 (NLSY79) is 0.76 to 0.80 (Borghans et al., 2011b). Friedman and Streicher (1985) estimate correlations between 0.65 and 0.82 for 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 See Borghans et al. (2011b) for information on the Stella Marris secondary school and the analysis described in the text.
Grit is a measure of persistence on tasks (Duckworth et al., 2007). Source: Borghans et al. (2011a).

Borghans et al. (2008) show that adults spend substantially more time answering questions on IQ tests when rewards are higher. Subjects high in Emotional Stability and Conscientiousness are less affected by rewards. Similarly, Segal (2008) shows that introducing cash incentives for performance on the coding speed test of the Armed Services Vocational Battery (ASVAB) increases performance substantially, particularly for men with lower levels of Conscientiousness.

2.4 Are There Stable Personality Traits?

Many have questioned whether personality traits exist. The publication of Walter Mischel’s 1968 book, *Personality and Assessment*, gave rise to a heated “personality-situation” debate within psychology, which pitted the social psychologists who favored situational factors as explaining behavior against those who considered stable personality traits as more consequential. Mischel argued that aspects of situations overshadow any effect of personality on
behavior.\footnote{This theme has been picked up in behavioral economics. See \cite{Thaler2008} for a recent discussion.}

A large body of evidence reviewed in \cite{Almlund2011} shows that stable personality traits exist and are predictive of many behaviors.\footnote{See the special issue of \textit{Journal of Research in Personality} (43), entitled “Personality and Assessment at Age 40” for a recent discussion.} An important paper by \cite{Epstein1979} presents compelling evidence that, averaging over tasks and situations at a point in time, people act in a predictable fashion with a high level of reliability ($R^2$ of 0.6-0.8) of average behavior (“measured personality”) across situations. The incentives in any situation also matter. Heritability studies show that measures of personality traits tend to be about 40%-60% heritable, suggesting that something tied to the person, rather than the situation, influences behavior \cite{Bouchard2001}.\footnote{Devlin et al. (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 \cite{Rutter2006} and an emerging literature on epigenetics.} Evidence in neuroscience suggests that expression of traits is related to regions of the brain \cite{Canli2006, DeYoung2010}.\footnote{See the evidence collected in \cite{Borghans2008} and \cite{Almlund2011} for a variety of other traits.}

\section{2.5 The Evolution of Personality Traits Over the Life Cycle}

Even though personality traits are relatively stable across situations, they are not set in stone. They change over the life cycle. Figure 3 shows that Conscientiousness tends to increase monotonically over the life cycle. Other traits change in different ways over the life cycle.\footnote{McArdle et al. (2000).} Crystallized intelligence tends to increase monotonically for most of the life cycle, whereas fluid intelligence tends to peak in early adulthood and then decline.\footnote{\cite{McArdle2000}.}

This evidence does not address whether these changes occur naturally (“ontogenic change”) or whether they are due to changes in the environments commonly experienced over the life cycle (“sociogenic change”). No evidence is available in the published literature on the distributions of these profiles over the life cycle. \cite{Almlund2011} review the evidence on
how parental investment and interventions promote changes in personality.

Figure 3: Cumulative Mean-Level Changes in Personality Across the Life Cycle

Note: Cumulative d values represent total lifetime change in units of standard deviations ("effect sizes").
Source: Figure taken from Roberts et al. (2006) and Roberts and Mroczek (2008). Reprinted with permission of the authors.

2.6 The Predictive Power of Personality

Table 2 shows that personality traits predict many later-life outcomes as strongly as measures of cognitive ability. Conscientiousness – the tendency to be perseverant and hardworking – stands out as the most predictive of the Big Five traits across many outcomes. Figure 4 presents for males correlations between the Big Five and educational attainment, adjusting and not adjusting for fluid and crystalized intelligence. Conscientiousness predicts educational attainment more than either of the facets of intelligence. Similar patterns appear for many other outcomes, including labor market performance, grades, and health.

A recurrent finding in the literature is that measured IQ is highly predictive of performance on complex tasks and jobs (Gottfredson, 1997). The importance of IQ increases with

---

43 Results are similar for women (see Almlund et al., 2011).
44 See Almlund et al. (2011), Borghans et al. (2008), and Roberts et al. (2007) for comprehensive reviews of the evidence.
job complexity, defined as the information processing requirements of the job: cognitive skills are more important for professors, scientists, and senior managers than for semi-skilled or unskilled laborers (Schmidt and Hunter 2004). In contrast, the importance of Conscientiousness does not vary much with job complexity (Barrick and Mount 1991), suggesting that it pertains to a wider spectrum of jobs.

The literature in economics establishes that the same bundle of traits has different productivity in different tasks. People also differ in their endowments of traits. These two features lead to sorting in the tasks people pursue in life and are a manifestation of the general principle of comparative advantage in the labor market and in life. (See Almlund et al. 2011; Borghans et al. 2008; Cattan 2012; Heckman et al., 2006, 2011.)

Achievement test scores are crude, low-dimensional summaries of high-dimensional vectors of traits operating in conjunction with effort. It is unlikely that these summaries capture the precise combinations of traits required for success in specific life tasks. The thrust of recent research in personality and economics is to isolate the traits that determine life outcomes and to understand how those diverse traits determine choices of tasks.

Most of the evidence in personality psychology is correlational. The reported correlations do not prove that personality traits cause higher educational attainment although it is consistent with it. For example, the reported pattern in Table 4 could arise if educational attainment increased Conscientiousness. We next present causal evidence.
3 Causal Evidence

3.1 Problems with Establishing Causality

Most studies in personality psychology do not address the question of causality, i.e., whether measured traits cause (rather than just predict) outcomes. Empirical associations are not a reliable basis for policy analysis. In this section, we discuss difficulties in establishing causality. We also summarize several studies that provide evidence that personality traits cause outcomes.

We introduce a simple framework to analyze the effect of traits on outcomes and how traits evolve over time.\footnote{This framework draws on Almlund et al. (2011).} Equation (1) shows how an outcome at age $a$, $T_a$, which is the

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 21 and 94. Source: \cite{Almlund2011} German Socio-Economic Panel (GSOEP), waves 2004-2008.
performance on a task, depends on cognition $C_a$, personality $P_a$, other acquired skills such as education and job training $K_a$, and the effort allocated to the task $e_{T_a}$:

$$T_a = \phi_a\left( C_a, P_a, K_a, e_{T_a} \right) \quad a = 1, \ldots, A. \quad (1)$$

Equation (2) shows how the effort allocated to task $T_a$ depends on cognition $C_a$, personality $P_a$, other acquired skills $K_a$, incentives $R_{T_a}$, and preferences $\Upsilon_a$:

$$e_{T_a} = \psi_{T_a}(C_a, P_a, K_a, R_{T_a}, \Upsilon_a) \quad (2)$$

The effort applied to a task is the outcome of a choice problem that depends on traits, preferences, and incentives, much like a supply equation in the standard theory of consumer choice. Preferences can be thought of as additional traits. Some psychological theories posit that people have limited effort that they can divide among different tasks (See, e.g., Baumeister and Tierney, 2011).

Equations (1) and (2) formalize the difficulty in establishing a causal relationship between outcomes and traits. Multiple traits, effort, and acquired skills generate performance in a given task. Many studies in psychology and economics do not control for these inputs and equate measurement of a set of outcomes with the trait the analyst is trying to measure. This practice can lead to a substantial bias in inference about any particular trait.

An additional point is that most studies assume a linear relationship between outcomes and traits. This practice is particularly problematic for measuring personality traits, where the effect of a trait on an outcome is not always linear or even monotonic. Too much of a

---

46 The empirical relationship between measured preference parameters and Big Five measures is weak. 47 Selecting measures and verifying them is part of the sometimes mysterious and inherently subjective process of “construct validity” in psychology. For a discussion, see Borghans et al. (2008).
good thing can be bad \left( \frac{\partial \phi_a}{\partial P_a} < 0 \text{ for } P_a > \bar{P} \text{ for threshold } \bar{P} \right)$. For example, extreme levels of traits are associated with psychopathologies. High levels of Conscientiousness are associated with Obsessive Compulsive Disorder, which hinders task performance \cite{Samuel and Widiger 2008}. Nonlinearities can also arise when traits and incentives interact, as in the analyses of Borghans et al. \cite{Borghans et al. 2008} and Segal \cite{Segal 2008} who show that people with different personality traits respond differently to incentives on tests.

The traits and other acquired skills evolve over time through investment and habituation. Equation \ref{eq:traits} shows that traits at age $a + 1$ are age-dependent functions of cognitive ability, personality traits, other acquired skills, and investment $I_a$ at age $a$. In this way, previous levels of traits and acquired skill affect current levels of traits and acquired skill. Equation \ref{eq:traits} formalizes the notion that the traits and skills governing performance at a point in time are themselves the outcome of investment and habituation:

$$
(C_{a+1}, P_{a+1}, K_{a+1}) = \eta_a(C_a, P_a, K_a, I_a^{\text{Investment and experience}}), \quad a = 1, \ldots, A. \quad (3)
$$

In conjunction with resource constraints, a “deeper” set of preference parameters at age $a$ may govern investment decisions and effort allocated to tasks.

### 3.2 Extreme Examples of Personality Change

Laboratory experiments and brain lesion studies provide some of the most compelling evidence that personality traits can change and that the change affects behaviors. The most famous example is that of Phineas Gage, a railway construction foreman whose head was impaled by a metal spike. Miraculously he retained his problem solving abilities, but he changed from being polite and dependable to being rude and unreliable. His personality change caused him to lose his job and alienate family members \cite{Damasio et al. 2005}. Lab-

\footnote{Formally, this occurs when $\frac{\partial^2 \psi_a}{\partial P_a \partial I_a} \neq 0$.}
Laboratory experiments show that expressed traits can be manipulated temporarily. Magnetic disruption of the left lateral prefrontal cortex can increase experimentally elicited discount rates (Figner et al., 2010) and nasal sprays of oxytocin increase trust (Kosfeld et al., 2005).

3.3 Evidence from the GED Testing Program

The GED is a standardized achievement test that serves as an alternative to a high school diploma. High school dropouts can take the seven-hour GED exam to certify that they have the “general knowledge” of a high school graduate. The test is widely used. The GED testing program currently produces 12% of high school certificates each year in the United States. We draw on the analysis of Heckman et al. (2012) and first present results for males. The GED program provides insight into the effects of personality traits on outcomes. GED recipients have the same cognitive ability as high school graduates, but differ in their personality traits.

Table 4 shows the correlations between GED scores and other achievement test scores. GED test scores are strongly correlated with scores on other standardized achievement tests. The correlations range from 0.61 with the General Aptitude Test Battery (GATB) to 0.88 with the Iowa Test of Educational Development, the progenitor of the GED.
Table 4: Validities of GED Test

<table>
<thead>
<tr>
<th>Test</th>
<th>Correlation</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armed Forces Qualification Test (AFQT)</td>
<td>0.75 - 0.79 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Iowa Test of Educational Development</td>
<td>0.88 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>American College Test (ACT)</td>
<td>0.80 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Adult Performance Level (APL) Survey</td>
<td>0.81 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>New York’s Degrees of Reading Power (DRP) Test</td>
<td>0.77 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Test of Adult Basic Education (TABE)</td>
<td>0.66-0.68†</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>General Aptitude Test Battery (GATB)</td>
<td>0.61-0.67†</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>National Adult Literacy Survey (NALS) factor</td>
<td>0.78 ‡</td>
<td>Baldwin (1995)</td>
</tr>
</tbody>
</table>

† Uses mean GED subtest scores
‡ Uses a general GED factor

GED recipients are smarter than other dropouts. Figure 5 shows the distributions of a factor extracted from the components of the Armed Services Vocational Aptitude Battery (ASVAB) for male high school dropouts, GED recipients, and high school graduates. The sample excludes people who attend post-secondary education. The distribution of the scores of GED recipients is much more like that of high school graduates than that of high school dropouts.

⁴⁹Similar results are found for females.
Figure 5: Cognitive ability by educational status

Source: Reproduced from [Heckman et al. (2011)](source), which uses data from the National Longitudinal Study of Youth 1979 (NLSY79). Notes: The distributions above represent cognitive ability factors estimated using a subset of the Armed Services Vocational Aptitude Battery (ASVAB) and educational attainment as laid out in [Hansen et al. (2004)](notes). The sample is restricted to the cross-sectional subsample for both males and females. Distributions show only those with no post-secondary educational attainment. The cognitive ability factors are normalized by gender to be mean zero standard deviation one.

If they have the same cognitive ability as high school graduates, then why do they drop out of high school? Success in school requires other traits. On a variety of other dimensions, GED recipients behave much more like other dropouts. Figure 6 shows measures of early adolescent drug use, crime, sex, and violence extracted from three data sources: National Longitudinal Survey of Youth 1979 (NLSY79), National Longitudinal Survey of Youth 1997 (NLSY97), and National Educational Longitudinal Survey (NELS). For discussion of these data sets, see [Heckman et al. (2012)](notes).

Male high school graduates perform better on all measures than high school dropouts or GED recipients. GED recipients are much more similar to dropouts, but in several cases are statistically significantly *more likely* to engage in risky behaviors than other dropouts. On no outcome measure in that figure are dropouts statistically significantly more likely to engage in risky behaviors compared to GED recipients. Figure 7 summarizes these adolescent behaviors using a single factor and shows that unlike the cognitive summary measures, the distribution of the noncognitive (personality) summary measure of GED recipients is much closer to that of dropouts than to that of high school graduates.

---

50See National Longitudinal Survey of Youth 1979 (NLSY79), National Longitudinal Survey of Youth 1997 (NLSY97), and National Educational Longitudinal Survey (NELS). For discussion of these data sets, see [Heckman et al. (2012)](notes).
Figure 6: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates

(a) Smoking and Drinking
(b) Sex and Violent Behavior
(c) Criminal Behavior

Sources: [Heckman et al. (2012) Chapter 3]. National Longitudinal Survey of Youth 1979, National Longitudinal Survey of Youth 1997, National Educational Longitudinal Survey. Notes: Minor crime includes vandalism, shoplifting, petty theft, fraud, holding or selling stolen goods. Major crime includes auto theft, breaking/entering private property, grand theft. Violent crime includes fighting, assault, aggravated assault. Tests of Significance: The estimates for GED recipients and high school graduates are statistically significantly different at the 5% level for all variables. The estimates for dropouts and high school graduates are statistically significantly different at the 5% level for all variables, except for “Minor Crime (NLSY79)” and “Drinks by 14 (NLSY97).” The estimates of “Smokes by 14 (NLSY97),” “Drinks by 14 (NLSY97),” and “Theft by 14 (NLSY97)” between GED recipients and dropouts are statistically significantly different at the 5% level.
Figure 7: Distribution of a Summary Measure of Noncognitive Ability by Education Group

Source: Reproduced from Heckman et al. (2011), which uses data from the National Longitudinal Study of Youth 1979 (NLSY79). Notes: The distributions above represent non-cognitive ability factors estimated using measures of early violent crime, minor crime, marijuana use, regular smoking, drinking, early sexual intercourse, and educational attainment as in Hansen et al. (2004). Sample restricted to the cross-sectional subsample for both males and females. Distributions show only those with no post-secondary educational attainment. The non-cognitive ability factors normalized to be mean zero standard deviation one.

The traits that cause GED recipients to drop out of high school manifest themselves in many other life outcomes. One potential benefit of the GED certificate is that it opens doors to post-secondary education. Figure 8 shows post-secondary educational attainment for GED recipients and high school graduates. About 40% of GED recipients enroll in a 2- or 4-year college. Nearly half drop out within the first year. Only 3-4% earn a BA or AA degree by age 40.
Figure 8: Post-Secondary Educational Attainment Across Education Groups Through Age 40 (NLSY79) - Males

GED recipients lack persistence in a variety of tasks in life. Figure 8 shows the survival rates in employment (overall), employment in a given job, marriage, and in the condition of not having been incarcerated. GED recipients tend to exit employment, become divorced, and enter jail at rates similar to those of high school dropouts, while high school graduates are much more persistent.
Figure 9: Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates

(a) Survival Rate in Employment

(b) Survival Rate in Staying on a Job

(c) Survival Rate in Marriage

(d) Survival Rate in Staying Out of Jail

Source: [Heckman et al. 2012](#) Chapter 4. National Longitudinal Survey of Youth 1979 (NLSY79), nationally representative cross sectional sample. Notes: The spell to first time being incarcerated begins in the first year that individuals exit school. Tests of Significance: The estimates for GED recipients and high school graduates are statistically significantly different at the 5% level for all but the 2nd year of “Survival Rate in Not Having Been Incarcerated.” The estimates for dropouts and high school graduates are statistically significantly different at the 5% level for all but the 2nd year of “Survival Rate in Not Having Been Incarcerated.” The estimates for dropouts and GED are statistically only significantly different at the 5% level for the 5th year of the ”Survival Rate in Marriage.”

Adjusting for their differences in cognitive ability, male GED recipients perform virtually the same as high school dropouts in the labor market. Figure 10 shows the hourly wages and annual earnings of male GED recipients and high school graduates compared to high school dropouts for different age groups. The first set of bars shows the outcomes after adjusting for age, race, and region of residence. The second set of bars shows the effects
after additionally adjusting for scores on the Armed Forces Qualifying Test (AFQT). The third set of bars shows the effects after additionally adjusting for standard measures of family background. GED recipients and high school graduates outperform dropouts in regressions that only adjust for age, race, and region of residence. After adjusting for cognitive ability, GED recipients are indistinguishable from dropouts, whereas high school graduates earn more and have higher hourly wages. Controlling for family background characteristics does not change the story.
Figure 10: Labor Market Outcomes Differences - By Age - NLSY79 - Males

(a) Annual Earnings

(b) Hourly Wage

Source: Heckman et al. (2012, Chapter 3). National Longitudinal Survey of Youth 1979. Controls: “Raw” – age, race, and region of residence; “Abil” – age, race, region of residence, and AFQT adjusted for schooling at time of test; “BG” – mother’s highest grade completed, urban status at age 14, family income in 1978, broken home status at age 14, south at age 14, AFQT, and factors based on adolescent behavioral measures, crime and school performance. Regressions exclude those reporting earning more than $300,000 or working more than 4,000 hours. Notes: All regressions allow for heteroskedastic errors and when appropriate clustering at the individual level.
Most of the patterns found for women parallel those found for men. However, there are some important differences. While female GED recipients share similar cognitive and personality traits as male GED recipients, their outcomes differ. After accounting for differences in cognitive ability, female GED recipients do not earn higher hourly wages than other dropouts, but unlike men they have higher annual earnings because they are more likely to participate in the labor force.

3.4 Evidence from The Perry Preschool Program and Other Interventions

Evidence from the Perry Preschool Program shows how personality traits can be changed in ways that produce beneficial lifetime outcomes. The Perry preschool Program enriched the lives of three- and four-year-old low-income, Black children with initial IQs below 85 at age 3.

Participants were taught social skills in a “plan-do-review” sequence where students planned a task, executed it, and then reviewed it with teachers and fellow students. They learned to work with others when problems arose. In addition, home visits promoted parent-child interactions. The program ended after two years of enrollment and both treatments and controls entered the same school. The program was evaluated by the method of random assignment.

The program did not improve IQ scores in a lasting way. Figure [1] shows that, by age ten, 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

---

51 See Heckman et al. (2012).
52 The increased labor supply response is largely due to female GED recipients who attain some post-secondary education or who have dropped out of high school due to pregnancy. See Heckman et al. (2012) for a full discussion of the evidence on the performance of GED recipients.
53 We draw on the analysis of Heckman et al. (2012).
54 Sylva (1997) describes the Perry program as a Vygotskian program fostering personality traits. Vygotsky developed a psychology of child development in structured social settings that emphasized development of social and personality skills.
intervention studies.

Figure 11: Perry Preschool Program: IQ, by Age and Treatment Group

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

Nevertheless, the program improved outcomes for both boys and girls, resulting in a statistically significant rate of return of around 6-10% per annum for both boys and girls. (See \cite{Heckman et al. 2010}.) These returns are above the post-World War II, pre-2008 meltdown in stock market returns to equity estimated to be 5.8% per annum\footnote{See \cite{DeLong and Magin 2009}.}

The Perry Preschool Program worked primarily through improving personality traits. Participants had better direct measures of personal behavior (a weighted average of “absences and truancies,” “lying and cheating,” “stealing,” and “swears or uses obscene words” measured by teachers in the elementary school years). Participants of both genders improved their “externalizing behavior,” a psychological construct related to Agreeableness and Conscientiousness. For girls, the program improved Openness to Experience (proxied by academic motivation). The program also improved scores on the California Achievement
Test (CAT). This evidence is consistent with the evidence presented in the previous section that shows that performance on achievement tests depends on personality traits.

Other studies are broadly consistent with the evidence from the Perry Preschool study. Analyses of data from Project STAR, a program that randomly assigned kindergartners and teachers to classes of different sizes, yields results similar to the Perry Program. Using data from Project STAR, Dee and West (2011) find that assignment to a small class is associated with positive changes in personality. In a follow-up analysis, Chetty et al. (2011) examine the Project STAR program and find that students placed in higher quality kindergarten classes—as measured by their peer’s average performance on a Stanford Achievement Test—had significantly higher earnings in early adulthood.

The curriculum of Promoting Alternative Thinking Strategies (PATHS) teaches self-control, emotional awareness, and social problem-solving skills and is aimed at elementary school children (see Bierman et al., 2010). A recent random-assignment, longitudinal study demonstrates that the PATHS curriculum reduces teacher and peer ratings of aggression, improves teacher and peer ratings of prosocial behavior, and improves teacher ratings of academic engagement. PATHS is an exemplar of school-based social and emotional learning (SEL) programs. A recent meta-analysis shows that the program improved grades by 0.33 standard deviations and achievement test scores by 0.27 standard deviations (Durlak et al., 2011). Likewise, several random assignment evaluations of Tools of the Mind, a preschool and early primary school curriculum targeting development of self-control, show that it improves classroom behavior as well as executive function, defined as higher-level cognitive skills including inhibitory control, working memory, and cognitive flexibility (Barnett et al., 2008, 2006; Bodrova and Leong, 2001, 2007; Diamond et al., 2007; Lillard and Else-Quest, 2006). Positive findings are reported for the Montessori preschool curriculum (Lillard and

---

56 Bierman et al. (2010).
57 Note however that the largest federal study to date on character education programs, including PATHS, failed to find evidence for improvements in behavior or academic performance (see Social and Character Development Research Consortium, 2010).
58 However, a more recent large-scale study (Farran and Wilson, 2011) does not find any effect of the program on self-regulation or literacy, language, and mathematics achievement.
Else-Quest (2006). Unlike the Perry study, these studies do not have long-term followups.

There is evidence that targeted intervention efforts can improve aspects of Conscientiousness. In contrast to the multi-faceted curricula described above, studies targeting improvement in this trait are designed to isolate a particular mechanism producing behavioral change. For instance, Rueda et al. (2005) designed a set of computer exercises to train attention in children between four and six years of age. Children in the intervention group improved in performance on computer tasks of attention relative to children who instead watched interactive videos for a comparable amount of time. Similarly, Stevens et al. (2008) designed a 6-week computerized intervention and showed that it can improve selective auditory attention (i.e., the ability to attend to a target auditory signal in the face of an irrelevant, distracting auditory signal). As is typical of much of the literature, all of these programs have only short-term follow-ups.

Several studies suggest that personality can be remediated in adolescence. Martins (2010) analyzes data from EPSIS, a program developed to improve student achievement of 13-15 year-olds in Portugal by increasing 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 was tailored to each participant’s individual skill deficit. Overall, the program was successful and cost-effective, decreasing grade retention by 10 percentage points.

Other life experiences, like employment, can improve personality. Gottschalk (2005) analyzes evidence from a randomized control trial that working at a job can improve Locus of Control, a trait related to Neuroticism that measures the extent to which individuals believe that they have control over their lives through self-motivation or self-determination as opposed to the extent that the environment controls their lives (Rotter, 1966). He uses data from the Self-Sufficiency Project (SSP) in which some welfare recipients were randomly offered substantial subsidies to work. The subsidy more than doubled the earnings of a minimum wage worker. People in the experimental group worked about 30% more hours.

The relationship between Locus of Control and the Big Five trait of Neuroticism is discussed in Almhund et al. (2011).
than those in the control group. After 36 months, those who received the subsidy were more likely to have an improved Locus of Control.

3.5 Additional Evidence

Studies that account for the endogeneity of investment and education provide further evidence of the causal effect of education and cognitive and personality traits on outcomes. \cite{Heckman2011} estimate a sequential model of education to study the effects of education on a variety of outcomes. Correcting for selection into education, they find that early cognitive and personality traits affect schooling choices, labor market outcomes, adult health, and social outcomes and that increasing education promotes beneficial labor market, health, and social outcomes.

\cite{Heckman2006} estimate a version of Equation (3) to analyze the effects of increases in education on measured cognition and personality measures.\footnote{They estimate the effect of schooling on Self-Esteem and Locus of Control, personality traits related to Neuroticism. The Rosenberg Self-Esteem Scale attempts to assess the degree of approval or disapproval of oneself \cite{Rosenberg1965}. The relationship between these measures and the Big Five traits of Neuroticism is discussed in \cite{Almlund2011}.} Controlling for the problem of reverse causality that schooling may be caused by traits, they find that schooling improves both personality and cognitive traits and that these traits, in turn, boost outcomes.\footnote{Both \cite{Heckman2011} and \cite{Heckman2006} use an identification strategy based on matching on proxies for unobserved traits that corrects for measurement error and the endogeneity of schooling.}

\cite{Cunha2010} estimate a model of the technology of skill formation using longitudinal data on the development of children with rich measures of parental investment and child traits. They control for the endogeneity of investment using shocks to family income along with other instruments. Their model is a version of Equation (3). Traits are self-productive and exhibit dynamic complementarity – current values of traits affect the evolution of future traits through direct and cross effects. A leading example of a cross effect is that more motivated children are more likely to learn. They estimate parameters that summarize how past personality traits affect future cognitive traits.

They find that self-productivity becomes stronger as children become older, for both...
cognitive and personality traits. The elasticity of substitution for cognitive inputs is \textit{smaller} later in life. This means that it is more difficult to compensate for the effects of adverse environments on cognitive endowments at later ages than it is at earlier ages. This finding is consistent with the high rank stability of cognition over ages past 10-12 reported in the literature. It also helps to explain the evidence on the ineffectiveness of cognitive remediation strategies for disadvantaged adolescents documented in \textit{Cunha et al. (2006); Knudsen et al. (2006)} and \textit{Cunha and Heckman (2007)}. 

Personality traits foster the development of cognition but not vice versa. It is equally easy at all stages of the child’s life cycle to compensate for early disadvantage in endowments using personality traits. (Elasticities of substitution are essentially the same at different stages of the life cycle.) The most effective adolescent interventions target personality traits.

4 Summary

This paper reviews recent evidence on the importance of personality in economic and social life. It shows that success in life depends on many traits, not just those measured by IQ, grades, and standardized achievements tests. Personality traits predict and \textit{cause} outcomes.

All psychological traits are measured by performance on tasks. Psychological traits have different productivities in different tasks. Performance on tasks depends on incentives and multiple traits, giving rise to a fundamental identification problem when measuring any single trait. This identification problem is empirically important even for measures of cognitive traits.

The importance of cognitive ability increases with the complexity of the task. Given their endowments of traits and the incentives they face, people sort into tasks in life in pursuit of their comparative advantage.

Traits are stable across situations, but their manifestation depends on incentives to apply

\textsuperscript{62} \textit{Cunha et al. (2006)} report that 16\% of the variation in educational attainment is explained by adolescent cognitive traits, 12\% is due to adolescent personality (socioemotional traits), and 15\% is due to measured parental investments.
effort in the situations where they are measured and also depends on other traits and skills. However, traits are not set in stone. They change over the life cycle and can be enhanced by education, parenting, and environment to different degrees at different ages.

Scores on achievement tests capture both cognitive and personality traits. Children who are more academically motivated and more open to experience learn more and have higher test scores. More motivated children also try harder on achievement tests.

The evidence in this paper should give pause to analysts and policy makers who rely solely on achievement tests to monitor school performance and school systems. Standardized achievement tests do not adequately capture many skills that matter in life. GED recipients perform about as well as high school graduates on achievement tests but perform much worse in many aspects of life because they lack important personality traits. Categorizing GED recipients as high school graduates misrepresents national statistics on educational attainment. The Perry Preschool Program improved the lives of its participants without increasing their IQ scores, demonstrating why it is problematic to focus curricula exclusively on improving cognitive test scores.

Monitoring school progress and creating programs to enhance skills requires a broader framework of measurement. Interventions that promote beneficial changes in personality have an important place in a portfolio of public policies to foster human development.

\[^{63}\text{See }\text{Heckman and LaFontaine} (2010).\]
References


42


dynamic hypotheses of Gf-Gc theory using longitudinal life-span data. Learning and In-
dividual Differences 12(1), 53–79.

with general educational development (ged) credentials. Technical Report FR-PRD-84-6,
Human Resources Research Organization, Alexandria, VA.

Individual Achievement Test with a sample of Saskatchewan schoolchildren. Canadian


National Center for Education Statistics (Various). Digest of Education Statistics. Institute
of Education Sciences, U.S. Department of Education. Washington, DC.


Nisbett, R. E. (2009, February). Intelligence and How to Get It: Why Schools and Cultures


R IQs from scores on the Raven’s Standard Progressive Matrices. Journal of Clinical
Psychology 47(2), 277–284.


