Hard Evidence on Soft Skills

James J. Heckman and Tim Kautz

IRP Lampman Lecture
Madison, WI
May 16, 2012
This draft, March 29, 2012
Introduction

- Widespread use of standardized achievement tests.
- But the traits that they measure are not well-understood.
- This paper summarizes recent evidence on
  a. what achievement tests capture and what they do not;
  b. how achievement tests relate to other measures of “cognitive ability” like IQ and grades;
  c. the important skills that achievement tests miss or mismeasure, and
  d. how much these other 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 research 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.
- Standardized achievement tests: designed to capture “general knowledge” produced in schools and through life experiences.
- Such knowledge 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.
- This practice is inherently circular.
A more relevant validity criterion is how well these tests predict meaningful outcomes, such as educational attainment, labor market success, crime, and health.
Findings from Recent Research on Personality and Economics

- 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.
Do Stable Traits Exist?

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

- 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 just the situation, affects behavior.
Identification Problems: How are Personality Traits Measured?

- Psychological traits are not directly observed.
- 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.
- 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.
Psychologists sometimes claim to circumvent this measurement issue by creating taxonomies of traits and by applying intuitive names to responses on questionnaires.

They are still rooted in task performance or behavior.

Responding to a questionnaire is a task.
Determinants of Task Performance

- Performance in most tasks depends on preferences, personality traits, cognitive ability, and incentives although the importance of each differs by task.
- This dependence creates the fundamental identification problem in measuring traits.
- Measured cognitive ability and measured personality depend on a constellation of factors.
- Multiple traits affect performance on cognitive tasks.
Predictive Power of These Measures is High

- Despite these qualifications, measures of personality traits predict meaningful life outcomes.
- Conscientiousness – the tendency to be organized, responsible, and hardworking—is the most widely predictive of the commonly used personality measures.
Causal Status of Measured Traits

- Most studies in psychology only report correlations between measured traits and outcomes 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.

- On average, Agreeableness and Conscientiousness tend to grow with age.
Establishing Causality of the Traits

- First, we show how an achievement test, the General Educational Development (GED) test, fails to capture important traits that affect success in life.
- 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.
Defining and Measuring Personality Traits:
History and Measurement of Cognitive Ability

- IQ scores are widely interpreted as measuring a stable trait.
- Achievement tests are typically validated using other achievement tests, IQ tests, and grades, rather than with tasks or outcomes that matter.
### 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. (2011)</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. (2011)</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. (2011)</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>
**Definitions:**  

† 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
Validating cognitive ability tests using other measures of cognitive ability is inherently circular.

A more relevant measure is how these tests predict outcomes that matter.
### Table 2: Predictive Validities in Outcomes that Matter
(Adjusted R-Squared)

<table>
<thead>
<tr>
<th>Males</th>
<th>IQ Sample</th>
<th>AFQT Sample</th>
<th>GPA Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Pers</td>
<td>Both</td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.07</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.07</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.12</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Adj, $R^2$ Cog, Personality  
IQ Sample: 0.07  
AFQT Sample: 0.17  
GPA Sample: 0.11

<table>
<thead>
<tr>
<th>Females</th>
<th>IQ Sample</th>
<th>AFQT Sample</th>
<th>GPA Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Pers</td>
<td>Both</td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>-0.00</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>-0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.10</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Adj, $R^2$ Cog, Personality  
IQ Sample: 0.10  
AFQT Sample: 0.15  
GPA Sample: 0.10
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 from 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 intercourse 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 responded 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.
**Personality Measures**

- Personality traits are manifested through thoughts, feelings, and behaviors, and therefore, must be inferred indirectly by some measure of performance on “tasks,” broadly defined.
- 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: The Big Five Domains and Their Facets**

<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>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, Persistence, 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>
</tbody>
</table>
### Table 3: The Big Five Domains and Their Facets

<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>Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<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), Straight-forwardness (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>
</tbody>
</table>
### Table 3: The Big Five Domains and Their Facets

<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>Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<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, 1992). 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).
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.

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.
Identification Problems in Measuring Traits

- To infer traits from behaviors requires standardizing for all of the other contributing factors that produce the observed behavior.
There are two primary issues.

First, behavior depends on incentives created by situations.

Different incentives elicit different amounts of effort on the tasks used to measure traits.

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.

Examples: IQ and Effort
Figures 1 and 2 show how the variance in the scores on two achievement tests.
**Figure 1:** Decomposing Achievement Tests and Grades into IQ and Personality [NLSY79]

Source: Borghans et al. (2011). Notes: Rotter was administered 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 et al. (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 over-sample.
Figure 2: Decomposing Achievement Tests and Grades into IQ and Personality [Stella Maris]

Source: Borghans et al. (2011).
The Predictive Power of Personality

- Table 2 shows that personality traits predict many later-life outcomes as strongly as measures of cognitive ability.
### Table 2: Predictive Validities in Outcomes that Matter
(Adjusted R-Squared)

<table>
<thead>
<tr>
<th></th>
<th>IQ Sample</th>
<th>AFQT Sample</th>
<th>GPA Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Pers</td>
<td>Both</td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.07</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.07</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.12</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Adj, R^2 Cog, Personality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>-0.00</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>-0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.10</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Adj, R^2 Cog, Personality</strong></td>
<td>0.10</td>
<td>0.15</td>
<td>0.10</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 from 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 intercourse 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.
Figure 3: Association of the Big Five and intelligence with years of completed schooling

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: Almlund et al. (2011) German Socio-Economic Panel (GSOEP), waves 2004-2008.
- IQ is highly predictive of performance on complex tasks and jobs (Gottfredson, 1997).
- The importance of IQ increases with 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 Evolution of Personality Traits Over the Life Cycle
Figure 4: 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.
Causal Evidence

Problems with Establishing Causality

- Most studies in personality psychology do not address the question of causality, i.e., do measured traits cause (rather than just predict) outcomes?
- Equation (1) shows how an outcome at age $a$, $T_a$, which is the performance on a task, depends on cognition $C_a$, personality $P_a$, other acquired skills $K_a$, and the effort allocated to the task $e_{T_a}$:

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

(1)
Equation (2) shows how the effort allocated on the task depends on cognition $C_a$, personality $P_a$, other acquired skills $K_a$, incentives $R_{Ta}$, and preferences $\gamma_a$.

Preferences can be thought of as additional traits:

$$
e_{Ta} = \psi_{Ta}(C_a, P_a, K_a, R_{Ta}, \gamma_a).$$

(2)
Equations (1) and (2) formalize the threats to establishing a causal relationship between outcomes and traits.

Multiple traits, effort, and acquired skills generate performance in a given task.
The traits and acquired skills evolve over time through investment and habituation.

Equation (3) shows that traits at age $a + 1$ are a function of cognitive ability, personality traits, other acquired skills, and investment $I_a$ at age $a$.

$$(C_{a+1}, P_{a+1}, K_{a+1}) = \eta_a(C_a, P_a, K_a, \underbrace{I_a}_{\text{Investment and experience}})$$

$a = 1, \ldots, A$. 

(3)
Extreme Examples of Personality Change
Evidence from the GED Testing Program

- The GED testing program currently produces 12% of high school certificates each year in the United States.
- Table 4 shows the correlations between GED scores and other achievement test scores.
### 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.
Figure 5: Cognitive ability by educational status

Source: Reproduced from Heckman et al. (2011), 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). 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.
Figure 6 shows measures of early adolescent drug use, crime, sex, and violence extracted from three data sources.
**Figure 6:** Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Smoking and Drinking

**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, aggravatated assault.
Figure 6: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Sex and Violent Behavior

Figure 6: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Criminal Behavior

Figure 7 summarizes these adolescent behaviors using a single factor and shows that unlike the cognitive summary measures the distribution of the noncognitive summary measure of GED recipients is much closer to that of dropouts.
**Figure 7:** Distribution of a Summary 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.
Figure 8: Post-Secondary Educational Attainment Across Education Groups Through Age 40 (NLSY79) - Males


Tests of Significance: The estimates for GED recipients and high school graduates are statistically significantly different at the 5% level for all but attainment of the A.A. degree.
Figure 9: Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates: Survival Rate in Employment
Figure 9: Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates: Survival Rate in Same Job
**Figure 9: Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates: Survival Rate in Marriage**

![Graph showing survival rates over years since start of spell for different educational attainment levels: Dropout, GED, and High School Graduates.](image-url)
Figure 9: Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates: Survival Rate in Not Having Been Incarcerated

Survival Rate in Non-Incarcerated State

Years Since Start of Spell

- Dropout
- GED
- HSG
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.”
Figure 10 shows the hourly wages and annual earnings of male GED recipients and high school graduates compared to high school dropouts from age 20 to age 40.
Figure 10: Labor Market Outcomes Differences - By Age - NLSY79 - Males: Annual Earnings
Figure 10: Labor Market Outcomes Differences - By Age - NLSY79 - Males: Hourly Wage

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.
Evidence from The Perry Preschool Program and Other Interventions

- 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.
Personality psychologists mainly focus on empirical associations between their measures of personality traits and a variety of life outcomes. Yet for policy purposes, it is important to know mechanisms of causation to explore the viability of alternative policies. We use economic theory to formalize the insights of personality psychology and to craft models that are useful for exploring the causal mechanisms that are needed for policy analysis.

We interpret personality as a strategy function for responding to life situations. Personality traits, along with other influences, produce measured personality as the output of personality strategy functions. We discuss how psychologists use measurements of the performance of persons on tasks or in taking actions to identify personality traits and cognitive traits. We discuss fundamental identification problems that arise in applying their procedures to infer traits.

Many economists, especially behavioral economists, are not convinced about the predictive validity, stability, or causal status of economic preference parameters or personality traits. They believe, instead, that the constraints and incentives in situations...
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).
Figure 12: Decompositions of Treatment Effects on Outcomes, Males

Notes: The total treatment effect is normalized to 100%. One-sided \( p \)-values are shown above each component in each outcome. “(+)” and “(-)” denote positive and negative total treatment effects. “CAT total” denotes California Achievement Test total score.
Figure 13: Decompositions of Treatment Effects on Outcomes, Females

Notes: The total treatment effect is normalized to 100%. One-sided $p$-values are shown above each component in each outcome. “(+)” and “(-)” denote positive and negative total treatment effects. “CAT total” denotes California Achievement Test total score.
Figure 14: Decompositions of Treatment Effects, Factor Scores versus MLE

Cognition    Externalizing Behavior    Academic Motivation    Other Factors

# of misdemeanor violent crimes, age 27 (-)
- **SCORE**: 0.099
- **MLE**: 0.020

# of felony arrests, age 27 (-)
- **SCORE**: 0.120
- **MLE**: 0.048

# of misdemeanor violent crimes, age 40 (-)
- **SCORE**: 0.066
- **MLE**: 0.050

# of felony arrests, age 40 (-)
- **SCORE**: 0.050
- **MLE**: 0.025

# of misdemeanor arrests, age 27 (-)
- **SCORE**: 0.071
- **MLE**: 0.091

# of felony arrests, age 27 (-)
- **SCORE**: 0.071
- **MLE**: 0.085

# of misdemeanor arrests, age 40 (-)
- **SCORE**: 0.136
- **MLE**: 0.132

# of felony arrests, age 40 (-)
- **SCORE**: 0.056
- **MLE**: 0.079

Heckman and Kautz

Hard Evidence on Soft Skills 68 / 72
Summary

- Success in life depends on a multiplicity of traits, not just those measured by IQ, grades, and standardized achievements tests.
- All psychological traits are measured through behavior or performance on a task.
- Performance on tasks depends on incentives and multiple traits, giving rise to a fundamental identification problem when measuring any one trait.
Different tasks require different bundles of cognitive and personality traits.

Given their endowments of traits and the incentives, people sort into tasks in life in pursuit of their comparative advantage.

Traits are stable across situations, but their manifestation depends on the incentives in any situation where they are measured as well as on other traits and skills.

Scores on achievement tests capture both cognitive and personality traits.
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 fully capture other 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 *bundle* of traits captured by scores on achievement tests does not accurately measure the diverse traits required for success in life.

- Interventions that promote beneficial changes in personality have an important place in a portfolio of public policies to foster human development.