Advances in Measuring Non-Cognitive Skills

Tim Kautz
Mathematica Policy Research

RISE Conference
June 17, 2015
This draft, June 13, 2016
We have come to focus on achievement test scores to assess students, teachers, and schools (and even countries).

Achievement tests miss important non-cognitive skills that can be shaped through interventions and required for success in school and beyond (Kautz, Heckman, Diris, ter Weel, and Borghans, 2014).

Key policy-makers and organizations are planning to measure non-cognitive skills at a large-scale but there are challenges.

Recent advances in measurement provide some solutions to these challenges relevant for international use.
1. Why measure non-cognitive skills?
2. Traditional approaches to measuring non-cognitive skills
3. Challenges and advances in measuring non-cognitive skills
1. Why measure non-cognitive skills?
There are three main reasons to measure non-cognitive skills

1. IQ and achievement tests miss important skills
2. Non-cognitive skills predict important long-term outcomes
3. Non-cognitive skills are malleable and can be improved through interventions
(1) IQ and Achievement tests miss non-cognitive skills

- We have come to place great emphasis on cognitive tests (Heckman and Kautz, 2014)
- But these tests are not all that predictive of later life outcomes
Figure 1: Validities of Cognitive Measures in Age-35 Labor Market Outcomes (Adjusted R-Squared)

(a) Males

A common measurement system is the “Big Five” (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism)

Conscientiousness – the tendency to be organized and complete tasks – is the most predictive across a wide variety of outcomes
Figure 2: Association of the Big Five and Intelligence with Years of Schooling in GSOEP

(a) Males
Personality may affect health-related behavior, such as smoking, diet, and exercise. For example, Hampson, Goldberg, Vogt, and Dubanoski (2007) find that high scores of teacher assessments of Extraversion, Agreeableness, and Conscientiousness during elementary school predict overall health behaviors during midlife (less smoking, more exercise, better self-rated health) and indirectly affect health through educational attainment. The effects that were statistically significant at the 5% level or less ranged from 0.06 for the effect of Extraversion on physical activity to 0.12 for the effect of Conscientiousness on self-reported health status. Both the initial level and the growth in hostility (a facet of Neuroticism) throughout elementary school predict cigarette, alcohol, and marijuana use in high school, and sociability (a trait related to Extraversion) predicts drinking but not smoking (Hampson, Tildesley, Andrews, Luyckx, and Mroczek, 2010). As Fig. 1.19 illustrates, Heckman, Stixrud and Urzua (2006) find that their personality factor affects the probability of daily smoking for males. The gradient is steepest at the high and low quantiles of the distribution.

Non-cognitive skills are malleable and can be improved through interventions

- IQ becomes relatively rank stable by age 10 while non-cognitive skills are more variable (Almlund, Duckworth, Heckman, and Kautz, 2011)
- Neuroscience shows that this malleability is associated with the slow development of the prefrontal cortex (Walsh, 2005)
- Consistent with review of the intervention literature (Kautz, Heckman, Diris, ter Weel, and Borghans, 2014)
  - Only interventions that started before age 3 had a long-term effect on IQ
  - Many interventions starting after age 3 have effectively improved outcomes by improving non-cognitive skills
  - Adolescent interventions that teach personality skills in the workplace (or specific context) are promising
2. Traditional approaches to measuring non-cognitive skills
Non-cognitive skills are typically measured using self reports

- Self-reports are a primary method of measuring non-cognitive skills
- The “Big Five” is a relatively well-accepted taxonomy
# Table 1: The Big Five Traits

## OCEAN

<table>
<thead>
<tr>
<th>Trait</th>
<th>Definition of Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. <strong>Openness to Experience</strong></td>
<td>The tendency to be open to new aesthetic, cultural, or intellectual experiences.</td>
</tr>
<tr>
<td>II. <strong>Conscientiousness</strong></td>
<td>The tendency to be organized, responsible, and hardworking.</td>
</tr>
<tr>
<td>III. <strong>Extraversion</strong></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>
</tr>
<tr>
<td>IV. <strong>Agreeableness</strong></td>
<td>The tendency to act in a cooperative, unselfish manner.</td>
</tr>
<tr>
<td>V. <strong>Neuroticism</strong></td>
<td>Neuroticism is a chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes.</td>
</tr>
</tbody>
</table>
I see myself as someone who tends to be lazy

Rating scale: 1 – “strongly agree”, 5 – “strongly disagree”
3. Challenges and advances in measuring non-cognitive skills
Basic measurement challenges

- All psychological measurements are based on performance on a task (Heckman and Kautz, 2012)
- An interpretive problem lies at the heart of any psychological measurement system for any particular trait
- It is necessary to standardize for incentives and the effects of other traits in performing a task
Figure 4: Determinants of Task Performance

Incentives → Effort → Task Performance

Effort → Non-Cognitive Skills → Task Performance

Effort → Cognitive Skills → Task Performance

Task Performance:
- Test Scores
- Self-Reports
- Other Behaviors
Figure 5: Decomposing Variance Explained for Achievement Tests and Grades into IQ and Character: Stella Maris Secondary School, Maastricht, Holland

Can boost IQ by 15 points by giving candies for correct answers — the Black/White gap in IQ in U.S

Segal (2012) shows that introducing performance-based cash incentives in a low-stakes administration of a measure of IQ increases performance substantially among roughly one-third of participants.
Reference bias is one challenge with self-reports

- Respondents rate themselves relative to their peers rather than the population at large
- Reference bias can be especially problematic if comparing across different contexts (e.g. between countries and schools)
- A form of a situation (peer group) affecting measurement
Figure 6: National Rank in Big Five Conscientiousness and Average Annual Hours Worked

Source: The Conscientiousness ranks come from Schmitt, Allik, McCrae, and Benet-Martínez (2007). These measures were taken in 2001 (Schmitt, 2002). The hours worked estimates come from Organisation of Economic Cooperation and Development (2001). Notes: Several countries are omitted due to lack of data.
Anchoring vignettes address reference bias

- Anchoring vignettes are an additional question that helps to standardize the situation
- Describe a situation and ask the respondent to rate performance
“Aline leaves her belongings in a mess, hates cleaning the house, and usually doesn’t complete her homework.”

How organized do you think Aline is?

(1) Not at all (2) A little (3) Moderately (4) Very much (5) Completely
Kyllonen and Bertling (2013) demonstrates that including anchoring vignettes changes a cross-country relationship between teacher support and achievement from -0.45 to 0.29 in PISA 2012.
Figure 7: Determinants of Task Performance

- Incentives
- Effort
- Non-Cognitive Skills
- Cognitive Skills
- Task Performance

Inputs:
- Test Scores
- Self-Reports
- Other Behaviors
Explore alternative measures available to schools

- Measures traditionally viewed as “outcomes” contain information on non-cognitive skills
- “Real-world” non-cognitive skill measures: grades, absences, credits earned, disciplinary infractions
- Highly predictive of later behavior
Table 2: Predictive Validity ($R^2$) from Ninth-Grade Measures on Various Outcomes

<table>
<thead>
<tr>
<th>Ninth-Grade Measure</th>
<th>Explore Test</th>
<th>GPA</th>
<th>Credits</th>
<th>Absences</th>
<th>Discipline</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Score (Grade 11)</td>
<td>0.78</td>
<td>0.22</td>
<td>0.05</td>
<td>0.10</td>
<td>0.02</td>
<td>0.79</td>
</tr>
<tr>
<td>GPA (Grade 11)</td>
<td>0.21</td>
<td>0.49</td>
<td>0.28</td>
<td>0.20</td>
<td>0.05</td>
<td>0.52</td>
</tr>
<tr>
<td>Absences (Grade 11)</td>
<td>0.09</td>
<td>0.22</td>
<td>0.12</td>
<td>0.35</td>
<td>0.03</td>
<td>0.39</td>
</tr>
<tr>
<td>Arrested within 4 Years</td>
<td>0.06</td>
<td>0.14</td>
<td>0.12</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>Grad HS within 5 Years</td>
<td>0.11</td>
<td>0.35</td>
<td>0.36</td>
<td>0.23</td>
<td>0.06</td>
<td>0.41</td>
</tr>
<tr>
<td>Enroll College within 6 Years</td>
<td>0.15</td>
<td>0.20</td>
<td>0.16</td>
<td>0.12</td>
<td>0.03</td>
<td>0.25</td>
</tr>
<tr>
<td>Grad College within 10 Years</td>
<td>0.17</td>
<td>0.17</td>
<td>0.07</td>
<td>0.09</td>
<td>0.01</td>
<td>0.23</td>
</tr>
</tbody>
</table>
Figure 8: Predictive Validity of Cognitive and Non-Cognitive Skill for High School Graduation
Conclusions

- Non-cognitive skills predict outcomes and are malleable
- There are challenges with implementing traditional measures of non-cognitive skills at scale
- Recent advances suggest some promising methods to address the challenges


