NBER WORKING PAPER SERIES

SOME CONTRIBUTIONS OF ECONOMICS TO THE STUDY OF PERSONALITY

James J. Heckman
Tomáš Jagelka
Timothy D. Kautz

Working Paper 26459
http://www.nber.org/papers/w26459

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2019

This research was supported in part by: NIH Grants R24AG048081 and R37HD065072, a seed grant from the briq Institute on Behavior and Inequality, many years of funding from the Spencer Foundation, a grant from FAIR (Center for Experimental Research on Fairness, Inequality and Rationality) at NHH Norwegian School of Economics, and in-kind sponsorship through partnerships from the Institute for Economic and Social Research at Jinan University and the TrygFonden Centre for Child Research at Aarhus University. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders, the official views of the National Institutes of Health, nor the views of the National Bureau of Economic Research.

Some Contributions of Economics to the Study of Personality
James J. Heckman, Tomáš Jagelka, and Timothy D. Kautz
NBER Working Paper No. 26459
November 2019
JEL No. C91,C93,D12,D9,D91

ABSTRACT

This paper synthesizes recent research in economics and psychology on the measurement and empirical importance of personality skills and preferences. They predict and cause important life outcomes such as wages, health, and longevity. Skills develop over the life cycle and can be enhanced by education, parenting, and environmental influences to different degrees at different ages. Economic analysis clarifies psychological studies by establishing that personality is measured by performance on tasks which depends on incentives and multiple skills. Identification of any single skill therefore requires isolation of confounding factors, accounting for measurement error using rich data and application of appropriate statistical techniques. Skills can be inferred not only by questionnaires and experiments but also from observed behavior. Economists advance the analysis of human differences by providing anchored measures of economic preferences and studying their links to personality and cognitive skills. Connecting the research from the two disciplines promotes understanding of the number and nature of skills and preferences required to characterize essential differences.

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

Tomáš Jagelka
Institute for Applied Microeconomics
University of Bonn
Regina-Pacis-Weg 3,D
Bonn, -- 53113
Germany
jagelka.tomas@gmail.com

Timothy D. Kautz
Mathematica Policy Research
600 Alexander Park Drive
Princeton, NJ 08540
tkautz@mathematica-mpr.com

Economics and psychology have long traditions investigating preferences that affect choices. Psychologist L. L. Thurstone (1927) created the modern framework for characterizing choices among discrete outcomes used in psychology and economics. Economists and psychologists investigate: (a) attitudes toward risk ("risk aversion"), (b) responses to ambiguous situations (ambiguity aversion), (c) valuation of future benefits relative to current benefits (time preference), (d) altruism, (e) positive and negative reciprocity, (f) trust, and (g) the ability to make rational decisions, which is closely related to intelligence (Choi et al., 2014). There are many other aspects of choice behavior that are studied in both standard neoclassical and behavioral economics, as well as psychology (for discussions, see Becker et al., 2012, Loewenstein et al., 2001 or Falk et al., 2016).

Economists and psychologists also assess life-relevant skills. The dominant paradigm in economics is the human capital model developed by Becker (1964). In that framework, human capital is usually envisaged as a homogenous substance that raises productivity in the workplace and in the home. It is a skill that is produced by families, schools, and firms and that is embodied in individuals. Variations in levels of human capital explain variations in earnings and in a variety of other lifetime outcomes such as health, employment, and wealth, to name only a few.

In the early days of human capital theory, there was a raging controversy associated with the "signaling hypothesis" (see Spence, 1972) that claimed that job market returns to schooling were only a return to cognitive ability, and not to any acquired skill. IQ and human capital were assumed to be measured on the same one-dimensional scale.

As economists developed a deeper understanding of the determinants of earnings, they discovered that multiple skills were required in various labor market and life tasks (see Roy, 1951, Mandelbrot, 1962 and Bowles and Gintis, 1976). These skills include conscientiousness, ability to work with others, and a variety of attributes that personality psychologists codified and analyzed. While the evidence was initially fragmentary, it was strongly suggestive and stimulated much further work. Relevant market skills were understood to be heterogenous in nature. Different labor market and life tasks require skills in different amounts and proportions. Earnings are rewards to multiple tasks in different markets (Mandelbrot, 1962). People differ in their endowments of skills.

Economists contribute to personality psychology by using psychological skills (primarily the Big Five and its relatives) to predict life outcomes (see for example Heckman et al., 2006 and Flinn et al., 2019) and to evaluate the outcomes of educational interventions (see Cunha and Heckman, 2008b; Cunha et al., 2010; Heckman et al., 2013; Alan and Ertac, 2018; Kosse et al., 2014). They enrich personality psychology by developing and applying rigorous methodologies for measurement and assessment.

Economists relate the psychologists' skills to economic preference parameters which are the fundamental drivers of decisions in economic theory (see Becker et al., 2012, Borghans et al., 2008, and Jagelka, 2018). Research on this topic is rapidly expanding in economics. An open question for both fields is the number and nature of preferences, personality types, and skill indicators required to adequately characterize differences in human potential.

This research has illuminated the information contained in the scores from achievement tests. Modern societies rely on written tests to sift and sort people, to evaluate students and schools, and to assess the performance of entire nations.[1] Achievement tests play a prominent role. The OECD actively promotes PISA tests. In the US, high school dropouts can take a 7-and-a-half hour achievement test—the General Educational Development (GED) exam—to certify that they are equivalent to high school graduates.[2]

Achievement tests were developed in the mid-twentieth century to measure a new concept—"*general knowledge*"—in an attempt to measure skills that are useful inside and outside of

[1]The Programme for International Student Assessment (PISA) evaluates student performance in math, science, and reading across countries, and its results attract a lot of media attention and influence policy. Scores from the year 2000 PISA test led Germany to reevaluate its educational system and introduce a variety of reforms (Grek, 2009). The creators of the original PISA tests called them literacy tests, not achievement tests, because PISA was designed to capture how knowledge can be applied to other contexts (OECD, 2013b). However, this was also the goal of the original achievement tests. We are unable to find any studies establishing that the original PISA measures were fundamentally different skills from those measured by achievement tests. Recently, however, PISA 2012 has added some component tests designed to capture aspects of non-cognitive skills including openness, locus of control, and motivation (OECD, 2013a).

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

the classroom.[3] Their developers thought that they had designed pencil-and-paper tests that would predict success in the labor market, in education, and in many other aspects of life. However, initially they did not validate their predictive power for important life outcomes outside the classroom.

Instead, validation of these tests is usually circular as we illustrate below in our discussion of the GED. Achievement tests are typically validated using IQ tests and grades, and not by their ability to predict important nonacademic life outcomes. The recent literature has conducted more meaningful evaluations of these tests.

Achievement test scores predict only a small fraction of the variance in later-life success. For example, adolescent achievement test scores explain at most 17% of the variability in later-life earnings.[4] Measurement error accounts for at most 30% of the remaining variability.[5]

Achievement tests do not adequately capture *non-cognitive or socioemotional skills*, a broad set of characteristics including preferences and personality. Examples include perseverance (sometimes called "grit"), conscientiousness (also called "grit"), self-control, trust, attentiveness, self-esteem and self-efficacy, resilience to adversity, openness to experience, empathy, humility, tolerance of diverse opinions, and the ability to engage productively in society, which are valued in the labor market, in school, and in society at large. Until recently these skills have largely been ignored in evaluations of persons, schools, and interventions to improve lifetime prospects. In recent research economists and psychologists have constructed measures of these skills and provide evidence that they are stable across situations and predict meaningful life outcomes.[6]

Skills are not set in stone at birth and determined solely by genes. They can be fostered.

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

[4]See Heckman and Kautz (2012). IQ tests alone explain at most 7% of this variability. Other work by Borghans et al. (2011b) shows that achievement tests explain a smaller proportion of lifetime success.

[5]See Bound et al. (2001).

[6]See the studies by Borghans et al. (2008) and Almlund et al. (2011). The modern literature traces back to Bowles and Gintis (1976), and Bowles et al. (2001). An important study in sociology is the work of Peter Mueser reported in Jencks (1979).

Cognitive and non-cognitive skills change with age and with instruction. Interventions to improve skills are effective to different degrees for different skills at different ages. Importantly, non-cognitive skills are more malleable at later ages than cognitive skills.

A growing body of empirical research shows that non-cognitive skills rival IQ in predicting educational attainment, labor market success, health, and criminality.[7] Both IQ and non-cognitive skills predict scores on achievement tests. Non-cognitive skills predict outcomes above and beyond their effects in predicting scores on achievement tests.[8]

This essay provides an overview of the main findings from the study of personality by economists. It acquaints psychologists with recently developed economic frameworks and the frontier of research in economics. It challenges psychologists to make sharper measurements with conceptually stronger foundations.

The rest of the essay is in three parts. Part 1 discusses issues of measurement. Part 2 presents evidence on the predictive power of personality skills. Part 3 discusses research on skill formation.

1 Measuring Cognitive and Non-cognitive Skills

1.1 Cognitive Skills

Measures of cognition have been developed and refined over the past century. Cognitive ability has multiple facets.[9] Psychologists distinguish between fluid intelligence (the rate at which people learn) and crystallized intelligence (acquired knowledge).[10] Achievement tests are designed to capture crystallized intelligence,[11] whereas IQ tests like Raven's progressive

[7] See Heckman and Kautz (2012, 2014a,b), Almlund et al. (2011), Borghans et al. (2008), and Roberts et al. (2007) for reviews.

[8] See Kautz and Zanoni (2019).

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

[10] See, e.g., Nisbett et al. (2012).

[11] Roberts et al. (2000).

matrices (1962) are designed to capture fluid intelligence.[12,13]

This new understanding of cognition is not widely appreciated. Many use IQ tests, standardized achievement tests, and even grades as interchangeable measures of "cognitive ability" or intelligence.[14] Scores on IQ tests and standardized achievement tests are strongly correlated with each other and with grades.[15] However, these general indicators of "cognition" measure different skills and capture different facets of cognitive ability.[16,17] Scores on these tests are also influenced by non-cognitive skills and the amount of effort that a test taker exerts.

1.2 Measuring Non-cognitive Skills

We use the term *non-cognitive skills* to describe the personal attributes not thought to be measured by IQ tests or achievement tests. These attributes go by many names in the literature, including soft skills, personality traits, non-cognitive abilities, character skills, and socio-emotional skills. These different names connote different properties.[18] "Traits" suggests a sense of permanence and possibly also of heritability. "Skills" suggests that these attributes can be learned. In reality, the extent to which these personal attributes can change lies on a spectrum. Both cognitive and non-cognitive skills can change and be changed over the life cycle, but through different mechanisms and with different ease at different ages. We use the term *skill* throughout this paper precisely because all attributes can be shaped.

Personality psychologists have studied noncognitive skills for the past century. Psycholo-

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

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

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

[15]See Heckman and Kautz (2012).

[16]See Borghans et al. (2011a).

[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 were once thought to be a more objective measure of student quality than high school grades (Lemann, 1999).

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

gists primarily measure noncognitive skills by using self-reported surveys or observer reports. They have arrived at a relatively well-accepted taxonomy of non-cognitive skills called the Big Five, well-known to readers of this volume. Some argue that the Big Five are the "longitude and latitude" of non-cognitive skills, by which all more narrowly defined skills may be categorized.[19]

While the Big Five measures are now widely used in psychology, there are several other taxonomies, including the Big Three, the MPQ, and the Big Nine that are conceptually and empirically related to the Big Five.[20] Other taxonomies, including psychopathology as measured by the DSM-IV and measures of temperament, have also been related to the Big Five.[21] In earlier research, Almlund et al. (2011) and Becker et al. (2012) summarize evidence showing that economic preference parameters are not closely related to the Big Five measures and apparently represent different attributes. Preference parameters along with the non-cognitive skills measured by personality psychologists govern behavior.[22] Recent work by Jagelka (2018) challenges this research and establishes a much tighter link than previously thought due to careful treatment of measurement error and mistakes made by persons being surveyed. The search is on for a minimal set of skills required to characterize empirically based human differences. The field is currently wide open.

1.3 A Task-Based Framework for Identifying and Measuring Skills

A leading personality psychologist (and author in this volume) defines personality (non-cognitive) traits (skills) as follows:

Personality traits are the relatively enduring patterns of thoughts, feelings, and

[19]Costa and McCrae (1992).

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

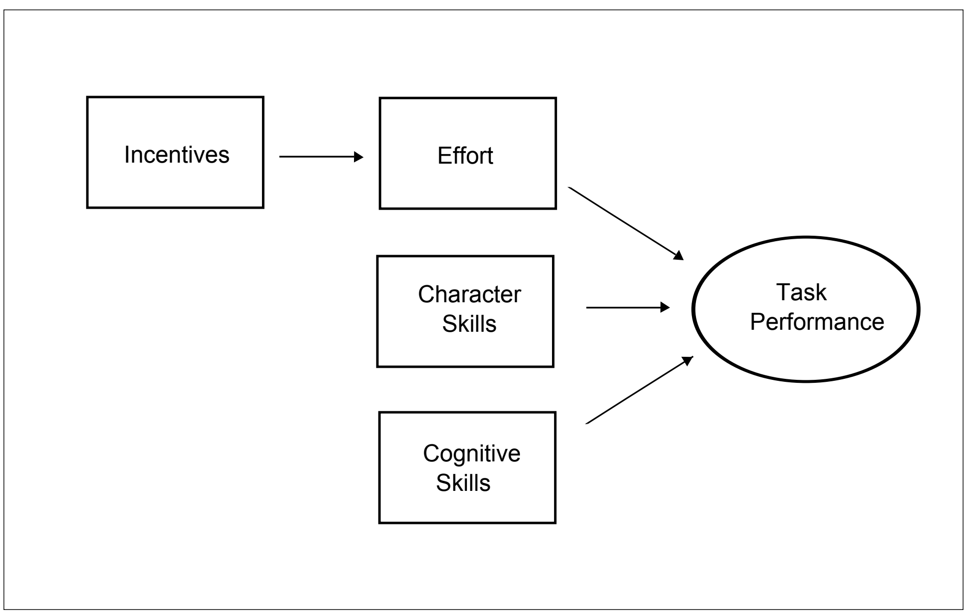
[21]See, e.g., Cloninger et al. (1999).

[22]A deeper question, 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 more basic set of preferences or goals. McAdams (2006) adds goals to the list of possible skills. Almlund et al. (2011) and Heckman and Kautz (2012) develop a model in which preferences and endowments of skills determine the effort applied to tasks. As shown in the next section, performance on tasks is the source of any measurement of a skill. Hence, in their framework, *measures* of skills are determined, in part, by preferences.

behaviors that reflect the tendency to respond in certain ways under certain circumstances. (Roberts, 2009, 140)

Roberts' definition of personality ("non-cognitive") skills, and the one favored by Almlund et al. (2011), suggests that all psychological measurements are calibrated on measured behaviors or "tasks" broadly defined. Tasks include taking IQ tests, answering personality questionnaires, performing a job, attending school, completing secondary school, participating in crime, or performing in an experiment run by a social scientist. Figure 1 depicts how performance on a task can depend on incentives, effort, and cognitive and non-cognitive skills. Performance on different tasks depends on these components to different degrees.[23] People can compensate for their shortfalls in one dimension by having strengths in other dimensions.

Figure 1 Determinants of Task Performance



Common practice in personality psychology assesses skills by self-reported questionnaires (see John, 2000; John and Srivastava, 1999a for one widely used questionnaire). However,

[23]There is a growing body of research which shows that performance on tasks is moderated by a variety of behavioral biases (see Chapman et al., 2018, and Stango and Zinman, 2019 for recent summaries).

performance on any task or any observed behavior can be used to measure personality and other skills.[24]

However, one needs to be careful in going this route. For example, completing high school requires many other skills besides those measured by achievement tests, including showing up in school, paying attention, and behaving in class.[25] A good score on an IQ test requires both ability and effort, which is affirmed by incentives to succeed in this task. While this point is generally known, it tends to be forgotten in naive use of personality and achievement inventories.

Inferring skills from performance on tasks requires standardizing all of the other contributing factors that produce the observed performance. The inability to parse and localize behaviors that depend on a single skill or ability gives rise to a fundamental problem of assessing the contribution of any particular skill to the successful performance on any task (or measure). This problem is commonly ignored in empirical research that studies how cognitive and non-cognitive skills affect outcomes.[26]

There are two distinct issues that need to be addressed in designing measures of skills based on performance of any task. First, behavior depends on incentives created by situations. Different incentives elicit different amounts of effort on the tasks used to measure skills. Accurately measuring non-cognitive skills requires standardizing for the effort applied in any task. Second, performance on most tasks depends on multiple skills. Not standardizing for other relevant skills that determine performance on a particular task used to measure a particular skill can produce misleading estimates of that skill.

These issues are empirically relevant. For example, incentives can increase effort and thus influence scores on IQ tests. Studies conducted over the past 40 years show that incentives

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

[25] The idea of using behaviors to measure non-cognitive is old. Ralph Tyler suggested using measures of behavior to capture non-cognitive skills in his first proposal for the National Assessment of Educational Progress tests. See Tyler (1973) and Rothstein et al. (2008). This idea is being been pursued in the recent literature (Heckman et al., 2018; Jackson, 2018). See Kautz and Zanoni (2019) for a recent application of this idea. We discuss this approach more extensively in Section 1.5.

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

can increase IQ scores, particularly among low-IQ individuals. Providing M&M candies for correct answers can increase scores by an amount equivalent to the black-white gap in IQ.[27] However, there is no evidence that this incentive-induced performance persists. It has yet to be shown that creating incentives for performance on one test improves performance on subsequent tests even of the same nature, or in any other life task. Indeed, there is some evidence that such incentives, in fact, may worsen subsequent performance (Deci and Ryan, 1985; Ryan and Deci, 2000).

Not all persons respond with equal strength to incentives. Research by Borghans et al. (2008) and Segal (2012) shows that the responsiveness of persons to incentives on IQ and achievement tests depends on their non-cognitive skills. Borghans et al. (2008) survey a body of literature that establishes the power of incentives to shape IQ scores. Duckworth et al. (2011) show that the motivation of test takers predicts IQ scores.

The same issues apply to measures of non-cognitive skills. While less is known about the degree to which situations or incentives can affect how people respond to self-reported measures of non-cognitive skills, Chen et al. (2019) provide some experimental evidence. They conduct two experiments that show how survey conditions—part of a student's situation—can affect student responses on the Big Five questionnaire. The first experiment shows that providing information about the Big Five affects how students report on their skills. The second experiment shows that providing incentives for performance on a separate task affects how students respond on the primary task. This evidence demonstrates the need to standardize for aspects of the situation and incentives when measuring non-cognitive skills.

The recent literature shows that non-cognitive skills predict standardized achievement test scores, which some psychologists assume are good measures of intelligence.[28] Figure 2 (based on Dutch data) shows how the variability across persons in the scores on one

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

[28] See, e.g., Nisbett (2009).

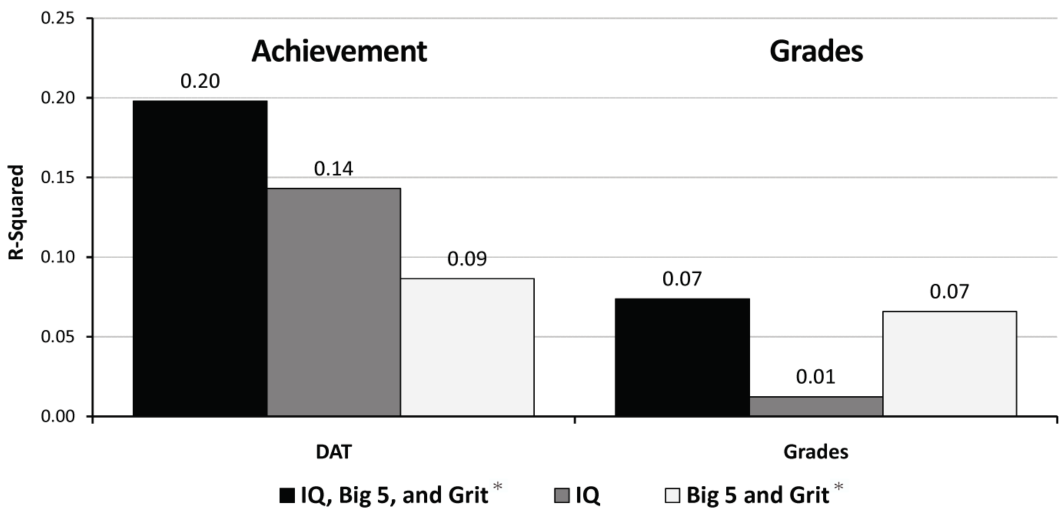
achievement test, the Differential Aptitudes Test (DAT),[29] are determined by IQ and non-cognitive measures. Non-cognitive skills explain a substantial portion of the variability across persons in DAT scores. Non-cognitive skills explain the variance in achievement scores above and beyond the variance that IQ explains when both measures of non-cognitive skill and IQ are included in a regression. These findings caution the interpretation that standardized achievement tests only measure cognitive ability. They also capture non-cognitive skills.[30]

The Bell Curve by Herrnstein and Murray argues that intelligence is a primary determinant of later-life outcomes. However, they use an achievement test (the Armed Forces Qualification Test, AFQT) as their measure of intelligence. Therefore, as discussed above, their findings implicitly show the power of both cognitive and non-cognitive skills in shaping life outcomes in the United States. Furthermore, recent studies have also demonstrated that other measures of non-cognitive skills can add predictive power above and beyond achievement tests (see Kautz and Zanoni, 2019; Heckman et al., 2014).

[29] The correlation between DAT and the widely used Armed Forces Qualification Test (AFQT) scores in the National Longitudinal Study of Youth 1979 (NLSY79) is 0.75 (Borghans et al., 2011c). Friedman and Streicher (1985) estimate correlations between 0.65 and 0.82 in a sample of high school sophomores and juniors. Kettner (1976) estimates correlations between DAT and the AFQT subtests of 0.76 to 0.89 in a sample of juniors and seniors.

[30] In the Stella Maris data, openness to experience is strongly correlated with IQ. See Borghans et al. (2011c).

Figure 2 Decomposing Variance Explained for Achievement Tests and Grades into IQ and Non-Cognitive Skills: Stella Maris Secondary School, Maastricht, Holland



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

1.4 Reference Bias

Answers from self-reports can be misleading when comparing levels of personality skills across different groups of people. Most personality assessments do not anchor their measurements in any objective outcome.[31] For example, the German Socio-Economic Panel (GSOEP) asks respondents to rate themselves on the following statement: *"I see myself as someone who tends to be lazy"* (Lang et al., 2011). The scale ranges from 1 = "strongly disagree" to 7 = "strongly agree." In answering this question, people must interpret the definition of *"lazy,"* which likely involves comparing themselves to other people. If different groups have different standards or reference points, comparing skills across groups can be highly misleading. Laziness may mean different things to different groups of people.

This measurement problem—sometimes called reference bias—is empirically important.[32] Schmitt et al. (2007) administer a Big Five personality questionnaire to groups of people

[31] These are called Likert scales (Likert, 1932).
[32] Reference bias is also problematic in health surveys that use self-reported, subjective health assessments. See Groot (2000).

in a variety of different countries. Using their estimates, Figure 3 shows how Organization of Economic Cooperation of Development (OECD) countries rank (from high to low) in conscientiousness – the tendency to be hard-working and persistent. The bars display the average number of hours that people work in the country.

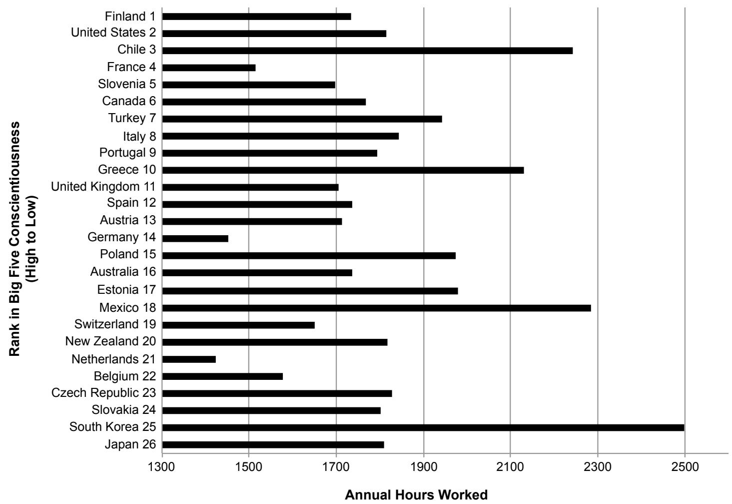
At first glance the results are surprising. South Korea ranks second to last in terms of conscientiousness but also ranks first in the number of hours worked. South Korea is not an anomaly. Country-level reports of Big Five conscientiousness are unrelated to the number of hours worked. The rank correlation between hours worked and conscientiousness across countries is negative, though statistically insignificant.[33] This finding contrasts with studies showing that non-cognitive skills tend to be positively related with labor supply within individual countries.[34]

Anchoring vignettes improve the performance of surveys. For example, the PISA 2012an instrument that assesses student achievement across countries also included a teacher support scale based on subjective student ratings of their teachers and corresponding anchoring vignettes. Before adjusting responses using vignettes, the correlation across countries of achievement and (positive) teacher support suggested a negative relationship between the two variables (Kyllonen and Bertling, 2013). This result is surprising since teacher support and achievement are typically positively related. After adjusting the scale using the vignettes, the correlation between the two variables increased and became positiveas would be expected. The unadjusted negative relationship was the misleading result of reference bias. In addition, Primi et al. (2016) found that anchoring vignettes improved internal consistency (reliability) for measures of socioemotional skills, that is, it improved the extent to which a group of items hung together.

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

[34] See the studies summarized in Almlund et al. (2011).

Figure 3 National Rank in Big Five conscientiousness and Average Annual Hours Worked



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

These findings are relevant for constructing measures of non-cognitive skills. Granting that achievement tests miss important skills, would the self-reported Big Five be a useful supplement to school evaluations? The possibility of reference bias suggests that it might not. For example, self-reports of conscientiousness might measure different things for different schools.

This issue is especially relevant given that some school districts are using self-reported measures of non-cognitive skills in their accountability systems (West, Buckley, Krachman, and Bookman, 2018). One obvious concern with accountability is that teachers or principals might coach students to give favorable responses on non-cognitive skill surveys. The evidence of Chen et al. (2019) previously discussed and the potential for reference bias suggests that such surveys are susceptible to more subtle biases.

Psychologists have addressed this problem.[35] Some surveys include vignette-based questions that attempt to standardize for aspects of the culture or situation. They attempt to

[35]For further discussion of reference bias, see Duckworth (2012); Goldammer (2010); Peng et al. (1997); Heine et al. (2001, 2008, 2002); Naumann and John (2011); and Schmitt et al. (2007).

frame questions so that the people in the survey answer within a common situation. However, this approach might not work well for evaluating schools, especially if teachers have incentives to coach children on taking these tests so that they score better and give answers perceived to be positive. Direct use of standard psychological measures can be problematic.[36]

1.5 Studies Measuring Skills Using Behaviors

Ralph Tyler, who pioneered the development of achievement tests, recognized their limitations.[37] He suggested using measures of behavior such as performance, participation in student activities, and other observations by teachers and school administrators to complement achievement tests when evaluating students and schools. Several recent papers demonstrate that this is a promising approach. Heckman, Pinto, and Savelyev (2013) show that teacher ratings of elementary schoolchildren's behaviors are strong predictors of adult outcomes and that early childhood interventions promote the non-cognitive skills measured by these ratings. Heckman et al. (2014) estimate the causal effect of cognitive and socio-emotional (non-cognitive) skills on a variety of outcomes. They measure socio-emotional (non-cognitive) skills using risky and reckless behaviors measured in the adolescent years.[38] They develop and apply methods to use high school grades to measure both cognitive and non-cognitive skills. They show that non-cognitive skills promote educational attainment, beneficial labor market outcomes, and health.

Jackson (2018) studies the effect of teachers on student cognitive and non-cognitive skills. In a fashion similar to Heckman et al. (2018), Jackson measures cognitive skills using achievement test scores, while measuring noncognitive skills using absences, suspensions, grades,

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

[37] See Heckman and Kautz (2014a).

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

and grade progression. These measures of non-cognitive skills predict adult outcomes with a strength similar to measures of cognitive ability. His measures of non-cognitive skills are commonly available from the administrative records of schools. Kautz and Zanoni (2019) use early measures of behavior in school to predict graduation and college attendance.

Similar to Ralph Tyler's suggestion of using participation in extracurricular activities to measure noncognitive skills, Lleras (2008) measures noncognitive skills in part by using tenth grade participation in sports, academic clubs, and fine arts activities. Participation in these activities predicts educational attainment 10 years later, even after controlling for cognitive ability as measured by achievement tests.

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

Hirschi and Gottfredson (1993) suggest that objective behavioral measures might be preferred to self-reports, partly because the act of filling out a survey requires some level of self-control. Answering survey questions is in itself a task that relies on skills beyond the ones targeted by the survey. Similarly, any self-reported questionnaire is also based on a behavior in two distinct ways. On the one hand, completing the questionnaire is a behavior that could be influenced by aspects of the situation, such as through incentives to respond in a particular way or through survey administration conditions. On the other hand, many questionnaires require respondents to reflect on past behavior, which could in turn depend on the situation or incentives they faced at the time. For example, one commonly-used Big Five item asks respondents the extent to which they agree with the statement, "I see myself as someone who tends to be lazy" (John and Srivastava, 1999b). The extent to which they might have behaved in a lazy way might depend on the incentives they faced to work hard at the time.

Some criticize this approach and argue that it is tautological to use measures of behavior to predict other behaviors even though the measures are taken early in life to predict later life behaviors.[39] However, as suggested by Figure 1, all tasks or behaviors can be used to infer a skill as long as the measurement accounts for other relevant skills and incentives of the situation in which the task is performed. In addition, many of the recent studies in economics use early measures of behaviors to predict behaviors in adulthood (see, e.g., Heckman et al., 2013, 2014, 2018). Self-reported scales should not be assumed to be more reliable than behaviors, although personality psychologists often assume so. The question is which measurements are most predictive and which can be implemented in practice. The literature suggests that there are objective measurements of non-cognitive skills that are not plagued by reference bias.

1.6 Measuring Economic Preferences

In economic theory, preferences for risk and time are considered to be fundamental drivers of a wide range of decisions. Economists assume that people make choices so as to maximize their well-being or "utility." Attitudes toward risk are commonly embodied by the "coefficient of risk aversion", a parameter that governs the shape of an individual's utility function. Risk preference affects decisions beyond those directly related to uncertainty (such as whether to buy insurance). Attitudes towards time are captured by discount rates. They represent the rate that compares the value of outcomes that occur today to the value of outcomes that occur in the future. More recently, economists have devoted increasing attention to an expanded list of preferences such as altruism or trust, and to behavioral biases which skew decisions.

Choices on incentivized tasks are one way to measure preferences.[40] Games and choice tasks have been developed to elicit each preference with careful attention devoted to ensuring

[39] See the discussion in Pratt and Cullen (2000) and Benda (2005).
[40] See Golsteyn and Schildberg-Hörisch (2017) for a review.

that they are incentive compatible.[41] For example attitudes toward risk are commonly elicited by having an individual make a series of choices between a lottery and an increasing fixed payment (see Harrison and Elisabet Rutström, 2008 for an overview). Initially, when the fixed payment is low, most individuals will choose the lottery. They will switch to the fixed payment once it becomes sufficiently attractive to them. The later out in the sequence of choices when they switch, the more they are willing to tolerate risk. These tasks can be augmented to test for various behavioral biases such as loss aversion (by making one of the options in the lottery a negative payment, see Fehr and Goette, 2007) or certainty premium (by replacing the fixed payment in each choice task with a second lottery, see Callen et al., 2014). Analogously, time preference can be elicited from a series of choices between a fixed payment today and an increasing series of payments in the future (Andersen et al., 2008), altruism is measured as the amount that an individual would gratuitously transfer to another person out of an endowment given to him by an experimenter (Forsythe et al., 1994), etc. Each experimental task is designed to engage a specific preference while minimizing the role of confounding preferences or skills on performance.

Inferring preferences from choices which involve precise monetary tradeoffs avoids the reference bias discussed in the previous sections. As individuals are paid according to their choices, their decisions should be informative of their true preferences[42] and can be converted to specific preference parameters if one is willing to impose certain assumptions regarding the underlying structure of the utility function.[43] These parameters can then be compared across schools, districts, and cultures.

When incentivized elicitation is impractical due to logistical difficulties or budgetary constraints, researchers sometimes use hypothetical questions that do not offer actual payoffs.

[41] A mechanism is said to be incentive compatible if it provides individuals with an incentive to truthfully and fully reveal their preferences. As of now, there is no existing theory regarding incentive compatibility in making hypothetical choices. However, there is some evidence that people have a preference for being honest and for being seen as honest (Abeler et al., 2019).

[42] There is evidence that the level of incentivization (stakes) also matters (see Holt and Laury, 2002).

[43] Harrison and Swarthout (2016) show how appropriately designed experimental elicitation can be used to empirically test competing theories of decision-making under risk which allow for various biases.

However, this approach can result in "hypothetical bias."[44] Some questionnaires only compare hypotheticals, as in contingent evaluation surveys that ask respondents to place a value on hypothetical choices. Others have respondents compare choices between hypothetical and incentivized measures, i.e., between stated and revealed preference (see Harrison, 2006).

One possible approach to avoid hypothetical bias is instrument calibration. It involves experimentally identifying a hypothetical question or task that yields preferences that most closely approximate those elicited in an incentivized setting. Falk et al. (2018) implement this approach to elicit and compare economic preferences around the world. Another approach, statistical calibration, relies on identifying differences between preferences elicited in an incentivized and non-incentivized setting and correcting for them.[45] It is best employed when these differences vary in a systematic way with observed characteristics. Special care needs to be taken when eliciting preferences from young children. Simple, clearly explained games and choice tasks work best in this setting.[46]

1.7 Are Economists' Preferences Psychologists' Personality Skills?

Economists and psychologists consider preferences and personality skills respectively to be person-specific determinants of behavior which are stable across situations. The natural question is whether they represent two different approaches to studying the same underlying concepts or whether they capture fundamentally separate aspects of human differences. In the latter case combining them together would yield a more complete model for explaining human behavior. Despite an intuitive link between certain preferences and personality skills – risk aversion and extraversion, time preference and conscientiousness, altruism and agreeableness – early attempts to empirically document the hypothesized relationships yielded mixed results (see e.g., Daly et al., 2009; Dohmen et al., 2010; Anderson et al., 2011, and

[44]Hypothetical bias refers to differences in response between settings in which consequences are hypothetical and real.

[45]The responses from a hypothetical setting are corrected for "hypothetical bias" so as to match those which would have been obtained in an incentivized setting (where choices would have had real consequences).

[46]Sutter et al. (2019) provides a comprehensive summary of this literature.

Almlund et al. (2011) which summarizes existing literature up to the point of publication). Indeed, Becker et al. (2012) conclude that while both preferences and personality skills predict behavior such as labor market success, health, and life satisfaction, they are only weakly related and thus likely capture separate underlying constructs.

Recent work by Andersson et al. (2018) and Jagelka (2018) advances this literature. They show that failure to find a tight mapping between preferences and personality may be an artifact of the inherent difficulty in relating two unobserved constructs, both of which are difficult to measure.[47] They demonstrate that carefully separating true preferences from noise in observed decisions and in answers to survey questions about personality is essential in advancing this line of research. Jagelka (2018) shows that four factors related to cognitive ability and three related to the Big Five personality skills explain up to 50% of the variation in both average preferences for risk and time and in individuals' capacity to make consistent rational choices. Variation among individuals in conscientiousness by itself explains 45% of heterogeneity in people's impatience and 10% of the variation in their risk aversion. Furthermore, attitude towards risk is linked to both extraversion and to cognitive ability while the latter also explains individual propensities to make mistakes (i.e., to pick options which they in fact do not prefer).[48]

Over the past decades, behavioral economists have documented numerous departures from decision-making implied by standard preferences and rationality assumptions of neoclassical economic theory. Individuals with lower cognitive ability are more likely to make

[47] Other research also suggests that preferences and personality are measured with a significant amount of error. For example, Beauchamp et al. (2017) find that adjusting for *measurement error* increases both the predictive power of risk preference in terms of observed outcomes and leads to higher measured correlations with non-cognitive skills. Castillo et al. (2018) find that only risk aversion estimates corrected for *decision error* predict life outcomes.

[48] Risk and time preferences are also related to cognitive ability; however, the relationship is likely weaker. Higher cognitive ability is associated with more patient behavior (see for example Dohmen et al., 2018 and Jagelka, 2018). The effect of cognitive skills on risk aversion is more controversial and there have been studies which found a positive, negative or no link (see Dohmen et al., 2018 for a recent summary of the literature). Andersson et al. (2018) suggests that estimated relationships may have been biased by features of the experimental design while Jagelka (2018) shows that explicitly modelling for noise in observed choices and survey responses flips the estimated link between cognitive ability and risk preference from negative to positive.

make inconsistent choices (Choi et al., 2014), do not seem to learn efficiently by updating their priors on probabilities which they place on outcomes when presented with new evidence (El-Gamal and Grether, 2012), are overconfident with respect to their absolute and relative performance (Chapman et al., 2018), and so on. The literature documents that people sometimes seem aware of these biases and are willing to pay to have their choices restrained so that they do not regret them later (Augenblick et al., 2015).[49]

The proliferation of documented patterns in deviations from rational decision-making has led to efforts to synthesize them in a general tractable model using behavioral summary statistics (Gabaix, 2014 and Chetty, 2015). Stango and Zinman (2019) measure 17 behavioral biases linked to financial decision making alongside standard economic preferences, cognitive skills, and non-cognitive skills. Most individuals in their sample display multiple behavioral biases with the average person exhibiting 10 out of 17. They show that biases are related to cognitive but not to non-cognitive skills. They are highly correlated with financial outcomes: a one standard deviation increase in their summary of bias parameters is associated with an approximately 30% decrease in subjective and objective measures of an individual's financial condition.

Dean and Ortoleva (2019) present evidence that some of these biases are strongly correlated with each other, but there are still multiple independent dimensions. These findings suggest that decisions are strongly influenced by a rich interplay of preferences, biases, and skills as well as the other factors depicted in Figure 1. Economists and psychologists have begun to understand the drivers of heterogeneity in decisions using disparate techniques. Evidence is mounting that there is an overlap between the economists' system of preferences and the psychologists' system of personality. Interestingly, standard economic preference parameters are mainly related to non-cognitive skills while the ability to effectively implement those preferences in practice (the quality of decision-making) is associated with cognitive ability. Further study examining a broader array of preferences and skills will yield a more

[49]Note that "bias" is in general measured against a maintained model of rational decision-making, which, in fact, may not be widely accepted.

complete mapping and ultimately a unified set of determinants of human behavior drawing from the strengths and insights of the two fields.

1.8 Are Non-Cognitive Skills Stable Across Situations?

Many have questioned whether there are stable non-cognitive skills, i.e., whether people exhibit the same non-cognitive skills across different situations at a fixed point. Walter Mischel's 1968 book, *Personality and Assessment*, gave rise to a heated "personality–situation" debate within psychology, which pitted social psychologists who favored situational factors as primary determinants of behavior against personality psychologists who considered stable personality (non-cognitive) traits (skills) as more consequential. Mischel argued that aspects of situations overshadow any effect of personality (non-cognitive) traits (skills) on behavior. Ironically, Mischel himself later demonstrated the stability and predictive power of non-cognitive skills over the life cycle (as measured by the performance of subjects in demonstrating self-control in early childhood) in his celebrated "marshmallow experiment."[50] Oblivious to this evidence, behavioral economists continue to echo Mischel's 1968 claim. (See, e.g., Thaler et al., 2008).

A large body of evidence reviewed in Almlund et al. (2011) shows that stable non-cognitive skills exist and are predictive of many behaviors.[51] An early paper by Epstein (1979) presents compelling evidence that, averaging over tasks and situations, people act in a predictable fashion with a high level of reliability of average behavior ("measured non-cognitive skills") across situations.[52] Some non-cognitive skills show predictable patterns over the course of a person's life. Todd and Zhang (2019) present evidence on the life cycle evolution of non-cognitive skills using Australian longitudinal data. People become more

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

[51] See the special issue of *Journal of Research in Personality* (2009) entitled "*Personality and Assessment at Age 40*" for a recent discussion.

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

conscientious with age and more agreeable, but less extraverted. Their causal study shows that schools foster these skills. Overall, many non-cognitive skills seem to stabilize around the age of 30, but agreeables increases thereafter. However, there is little longitudinal evidence on the evolution of time and social preferences. See Golsteyn and Schildberg-Hörisch (2017) for a recent summary of the literature on the stability of preferences and personality.

2 The Predictive Power of Non-Cognitive Skills

2.1 Correlational Evidence

A substantial body of evidence shows that non-cognitive skills predict a wide range of life outcomes, including educational achievement, labor market outcomes, health, and criminality. For many outcomes, the predictive power of non-cognitive skills rivals that of measures of cognitive ability. Of the Big Five, conscientiousness – the tendency to be organized, responsible, and hardworking – is the most widely predictive across a variety of outcomes (see Borghans et al., 2008; Roberts et al., 2007; Almlund et al., 2011; Heckman and Kautz, 2012). Conscientiousness predicts years of schooling with the same strength as measures of intelligence (Almlund et al., 2011).

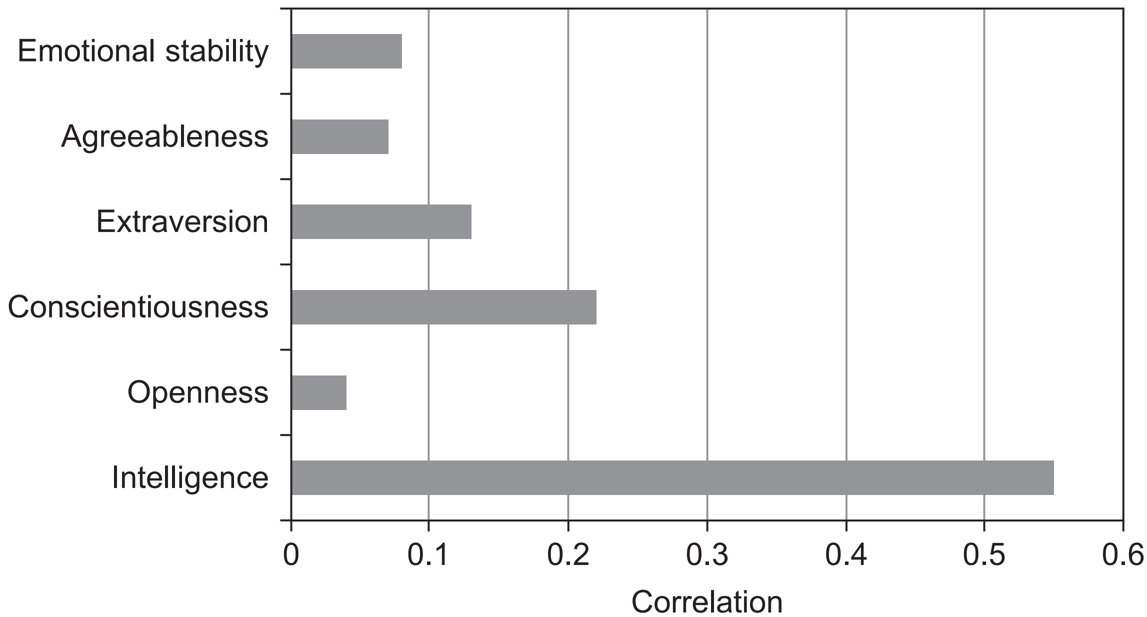
Aspects of job performance are also related to academic performance. Both require completing work on a schedule and involve intelligence to varying degrees. As with academic performance, numerous studies and meta-analyses have found that conscientiousness is associated with job performance and wages (Nyhus and Pons, 2005; Salgado, 1997; Hogan and Holland, 2003; Barrick and Mount, 1991). Figure 4 presents correlations of the Big Five and IQ with job performance. Of the Big Five factors, conscientiousness is the most strongly associated with job performance but is about half as predictive as IQ. Conscientiousness, however, may play a more ubiquitous role than IQ. The importance of IQ increases with job complexity (the information processing requirements of the job). Cognitive skills are more important for professors, scientists, and senior managers than for semiskilled or unskilled laborers (Schmidt and Hunter, 2004). The importance of conscientiousness does not vary much with job complexity (Barrick and Mount, 1991); this suggests that it applies to a wider spectrum of jobs.

Recently, other measures of persistence and self-control have been advocated as new and relevant dimensions of human potential. Grit is one of the proposed alternatives (see Duckworth et al., 2007 and Duckworth and Quinn, 2009). However, the case for grit as an independent dimension of human performance is, at best, weak. Credé et al. (2017) show that grit has negligible predictive power once one controls for conscientiousness.

Cobb-Clark et al. (2019) show in German data that self-control measured using the "Tangney scale" (Baumeister et al., 2004) predicts better health, educational, financial, and labor-market outcomes as well as increased self-reported life satisfaction. Using a series of natural experiments, they show that self-control is a cause rather than a consequence of increased schooling and that it is responsive to the overall institutional environment. Individuals who went to high school in East Germany prior to the fall of the Berlin Wall exhibit higher levels of self-control that those who lived in West Germany. After adding controls for cognitive ability, the Big Five, schooling, and economic preferences, the association between self-control and labor market outcomes goes away but remains statistically significant in explaining years of schooling, health, and life satisfaction. The authors thus conclude that beneficial effects of self-control are largely mediated by its impact on educational attainment. However, it is also possible that the marginal predictive power of self-control over economic preferences and the Big Five in this study can be explained by a more precise measure of self-control available in their data relative to other non-cognitive skills.[53]

[53] The authors use 13 indicators to construct a measure of self-control, but only 3 for each of the Big Five personality traits, and only one general question for measuring each of risk and time preference.

Figure 4 Associations with Job Performance



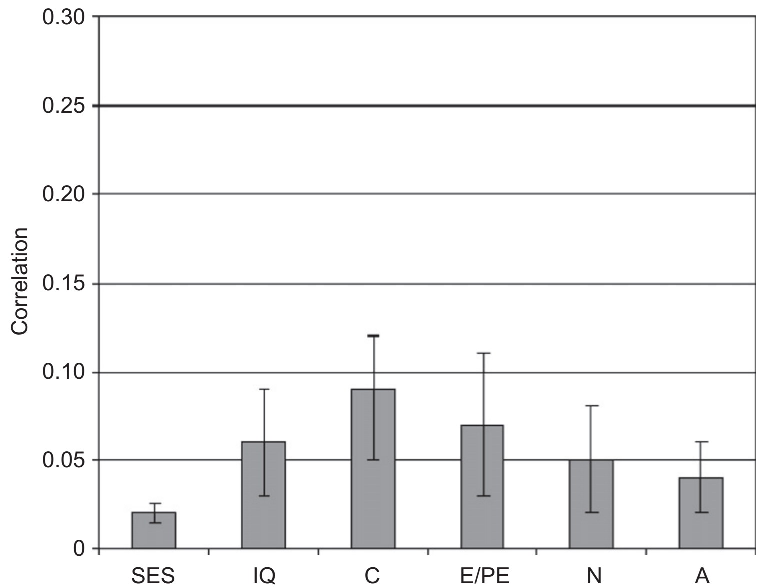
Source: The correlations reported for personality skills come from a meta-analysis conducted by Barrick and Mount (1991). The correlation reported for intelligence comes from Schmidt and Hunter (2004).
Notes: The values for personality are correlations that were corrected for sampling error, censoring,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

Measures of non-cognitive skills rival IQ and measures of socioeconomic status in predicting longevity.[54] Roberts et al. (2007) review evidence from 34 different studies on the predictive validity of the Big Five personality measures, relative to that of cognitive ability and socioeconomic status. Most studies they survey control for relevant background factors, including gender and severity of disease. Roberts and colleagues convert the results of each study into correlation coefficients that can be compared across studies. Figure 5 presents results from their analyses. Conscientiousness is a stronger predictor of longevity than any other Big Five skill and a stronger predictor than either IQ or socioeconomic status.[55] In general, skills related to conscientiousness, openness to experience, and agreeableness are

[54]For a recent study, see Friedman and Martin (2011).
[55]The timing of the measurements of non-cognitive skill relative to the outcomes varies by study.

associated with longer lives.[56] The magnitudes of the relationships vary across studies.

Figure 5 Correlations of Mortality with Non-Cognitive Skills, IQ, and Socioeconomic Status (SES)



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

Of the Big Five, conscientiousness and agreeableness are most predictive of criminality. In a sample of at-risk youth, boys who committed severe delinquent behaviors were more than three quarters of a standard deviation lower in agreeableness and conscientiousness, as measured by mothers' reports at age 12 or 13, than boys who had minor or no delinquent behaviors up to that age (John et al., 1994). Non-cognitive skills also predict financial outcomes. For example, households with a high internal locus of control[57] save more of their income (Cobb-Clark et al., 2016).

[56]See Friedman and Martin (2011); Martin et al. (2007); Kern and Friedman (2008); Mroczek and Spiro (2007); Boyle et al. (2005); Schulz et al. (1996); Kubzansky et al. (2001) .

[57]Judge et al. (2002) find that locus of control is highly related to emotional stability which is the inverse of neuroticism. Indeed, they say that "measures purporting to assess self-esteem, locus of control, neuroticism, and generalized self-efficacy may be markers of the same higher order concept."

There is a growing literature in economics on the predictive power of economic preferences and behavioral biases. Falk et al. (2018) conduct a comprehensive survey-based study of economic preferences across the world. They find that contemporaneous measures of preferences within countries are correlated with savings decisions, labor market choices, and prosocial behaviors. Furthermore, they present some evidence that differences in preferences across countries are related to differences in per capita income, entrepreneurial activities, and the frequency of armed conflicts. Choi et al. (2014) show that a 1 standard deviation increase in choice inconsistency is associated with an almost 20% increase in household wealth. Stango and Zinman (2019) find a similar impact of an increase in the number of exhibited behavioral biases on objective and subjective measures of financial well-being.

There is comparatively little evidence on the predictive power of early measures of preferences on later life outcomes. Golsteyn et al. (2014) show that survey-based time preferences elicited in Swedish data at age 13 predict outcomes decades later, mirroring the findings from Mischel's marshmallow study. In fact the impact seems to grow over time. Individuals who were more patient at age 13 had higher grades through high school and college, higher incomes, a lower probability of being unemployed, lower obesity, and a lower likelihood of dying before the age of 50. The effects are heterogenous and the association between early measures of patience and outcomes seems to be mediated by increased investments in human capital. Cadena and Keys (2015) confirm the link between patience and human capital accumulation using U.S. data. They find that impatience is related to time-inconsistent behavior such as dropping out of college after having already attended it for 3 years and conjecture that present bias (or short-run impatience) is the main culprit. Not only do impatient individuals accumulate less human capital and thus earn less, they also tend to express more regret later in life.

Accounting for randomness in observed decisions seems crucial and offers a promising avenue for future research. A large body of previous literature treated estimates of preferences and skills as if they were error-free. This is a strong assumption. Beauchamp et al. (2017)

show in a study of twins that accounting for measurement error increases the predictive power of estimates of risk preferences. Lower risk aversion is related to more risky financial decisions, a higher probability of owning a business but also higher alcohol consumption and more smoking. Castillo et al. (2018) find that after correcting for decision error, more risk averse 8th graders in Georgia, USA have fewer behavioral problems a year later and are more likely to graduate from high school.[58]

As with most studies in personality psychology, the evidence presented in Figures 4–5 and most of the literature do not address the question of causality; that is, do measured skills *cause* (rather than just predict) outcomes? Empirical associations are not a reliable basis for policy analysis. As previously noted (see Figure 1), multiple skills and effort all generate performance in a given task. Many studies in personality psychology do not control for all of the factors that produce performance on measured tasks. They equate measures of outcomes with the skill being measured.[59] This practice can lead to a substantial bias in inference about the importance of any particular skill. The discussion of the GED program and survey of the intervention literature in the section 3 presents evidence on the causal relationship between skills and outcomes.

2.2 The Skills Needed for Success in the Labor Market

Another perspective on the importance of non-cognitive skills comes from surveys of employers and workers. In a 1991 American report, the Secretary's Commission on Achieving Necessary Skills (SCANS) conducted an extensive analysis of which skills workers needed in the American workforce.[60] The Commission researched the literature, consulted with experts, and conducted detailed interviews with workers and/or supervisors in 50 occupations. The interviews rated the importance of various skills in the context of illustrative tasks and

[58] Castillo et al. (2011) showed that higher patience was also associated with fewer disciplinary referrals at school.

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

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

tools on the job. Using these sources, the Commission categorized necessary skills into basic skills, thinking skills, personal qualities, and a set of workplace competencies. In addition to reading, writing, and math skills, basic skills include listening and speaking. The thinking skills cover creative thinking, decision making, problem solving, reasoning, and the ability to learn. SCANS specifies that personal qualities include responsibility, self-esteem, sociability, self-management, integrity, and honesty. SCANS identifies five groups of workplace competencies: the ability to allocate resources (time, money, facilities), interpersonal skills (such as teamwork, teaching others, leadership), the ability to acquire and to use information, the ability to understand systems, and the ability to work well with technology.

Furthermore, some authors claim that the importance of social skills for success in the labor market is increasing. Deming (2017) claims that between 1980 and 2012, jobs requiring high levels of social interaction grew by nearly 12 percentage points as a share of the U.S. labor force. Social skills are complementary to cognitive skills and jobs which require a combination of both have become increasingly prevalent. Given the increased automatizing of the production process, social skills are at the heart of the remaining advantage of humans over machines. They allow workers to coordinate tasks more efficiently and thus to specialize in those tasks in which they are relative more productive. Accordingly, the labor market return to social skills was much greater in the 2000s than in the mid-1980s and 1990s while that of cognitive skills has declined slightly. However, Deming's estimates have been challenged and appear to be sensitive to alternative treatments of the same data. Caines et al. (2017a) show that the greatest growth in economic returns accrue to bundles of cognitive and non-cognitive skills and not to either separately. This is consistent with the pioneering work of Mandelbrot (1962).

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

Non-cognitive skills are especially critical for entry level and hourly workers. Of employers drawn from a national sample in the United States in 1996, 69% reported rejecting hourly applicants because they lacked basic employability skills, such as showing up every day, coming to work on time, and having a strong work ethic. This percentage is more than double the percentage of rejecting applicants due to inadequate reading and writing skills. Rejections for not passing a drug test were almost as common as rejections for lack of literacy skills.[61] In a 2007 survey of employers in Washington State, about 60% reported difficulty in hiring. They experienced less difficulty finding workers with adequate reading, writing, and math skills than with appropriate occupational, problem solving, teamwork, communication, and adaptability skills as well as positive work habits and a willingness to accept supervision.[62]

Evidence from the United Kingdom supports these findings. A 1998 survey of 4,000 employers found that the four skills found most lacking in 16 to 24-year-olds were technical and practical skills, general communication skills, customer handling skills, and teamwork skills.[63] At the bottom of the list were numeracy and literacy skills. In a 2002 survey of 4,000 employers in the UK, 23% of employers reported a significant number of their staff were less than fully proficient at their jobs. Skill shortfalls were most common in communication, teamwork, other technical and practical skills, customer handling, and problem solving and least common in numeracy and literacy.[64]

Consistent with these findings, the Confederation of British Industry defines employability as (1) values and attitudes compatible with the work, including a desire to learn, to

[61] Barton (2006).
[62] Washington Workforce Training Board (2008).
[63] Westwood (2004).
[64] Hillage et al. (2002).

apply that learning, to improve, and to take advantage of change; (2) basic skills (literacy and numeracy); (3) key skills (communication, application of numbers, information technology, improving one's own learning and performance, working with others, problem solving) sufficient for the needs of the work; (4) other generic skills such as modern language and customer service skills; and (5) job-specific skills and the ability to manage one's own career.

An ethnographic approach provides some revealing examples of how skills are used in context and how nonacademic skills are often developed and used as part of a "community of practice".[65] In addition to formal knowledge, Nelsen (1997) points out that workplaces require facts, principles, theories, and math and writing skills, but also informal knowledge embodied in heuristics, work styles, and contextualized understanding of tools and techniques. In her revealing case study of auto repair workers, Nelsen argues that social skills of new workers are very important for learning the informal knowledge of experienced workers, as captured in stories, advice, and guided practice. For a more recent discussion, see García (2014). The Atlanta Federal Reserve (Jones, 2018) recently issued a comprehensive assessment of workforce improvement strategies that include mentoring and fostering social skills of prospective workers. This reflects the growing understanding of the importance of multiple skills in the labor market and highlights the need to measure them appropriately.

Differences in personality and preferences appear to be responsible for some of the apparent puzzles concerning wage disparities between individuals and groups of individuals. Using Australian data, Cobb-Clark and Tan (2011) show that sorting into occupations by gender is in part driven by different sector-specific returns to non-cognitive skills for men and women. However, they find that the wage gap is largely driven by disparities for men and women within occupations.[66] Caliendo et al. (2015) explicitly model for the job search process and allow it to depend on an individual's locus of control and other characteristics. Using data from a German survey of the newly-unemployed, they find that people with a higher locus of control believe that effort in searching for a job is more productive. Accordingly, they

[65] Stasz (2001).

[66] Antecol and A.Cobb-Clark (2013) corroborate these findings using US data.

search more intensely and have higher reservation wages[67] than those with a lower locus of control. Flinn et al. (2019) build on this research, use the same dataset, and add wage bargaining which turns out to be important in explaining the gender wage gap. They show that heterogeneity in endowments of non-cognitive skills and in their valuations on the labor market largely account for the much discussed disparity. Specifically, women tend to score higher on agreeableness than men and are penalized for it in the wage bargaining process. Furthermore, men are compensated more highly for conscientiousness which is an important contributor to productivity.

2.3 Evidence from the GED Testing Program

Studies of the GED testing program provide additional causal evidence on the effect of non-cognitive skills on life-relevant outcomes. 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 was widely used but is now in decline. The GED testing program produced as many as 12% of high school certificates issued each year in the United States. The GED program provides insight into the effects of personality skills on outcomes. GED recipients have the same cognitive skill as high school graduates, but differ in their non-cognitive skills (see Heckman et al., 2014).

Table 1 displays the "validity" of the GED test as analyzed by psychometricians. It gives 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. By the standards of psychometrics, the GED test is "valid."

[67] A reservation wage is the minimum wage for which a person is willing to accept a job.

Table 1 Validities of GED Test

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

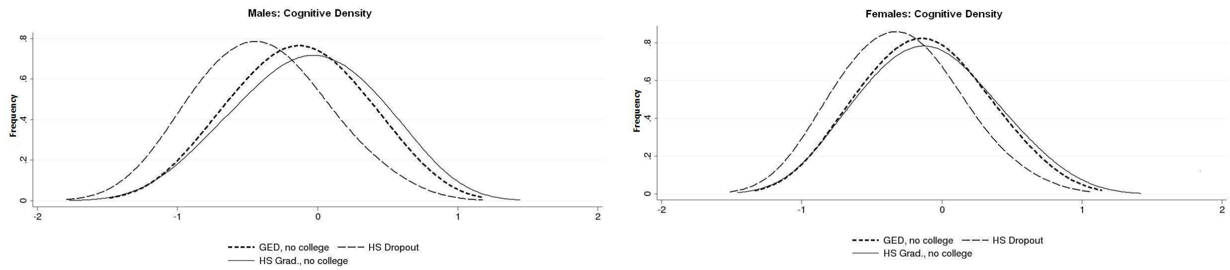
† Uses mean GED subtest scores
‡ Uses a general GED factor

By psychometric standards, GED recipients have higher skills than other dropouts. Figures 6a and b 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.[68] 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.

[68]Similar results are found for females.

Figure 6 Cognitive ability by educational status

(a) Males **(b)** Females



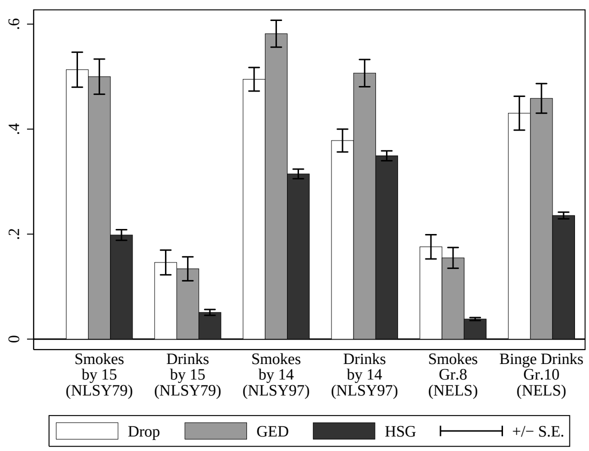
Source: Reproduced from Heckman et al. (2011), which uses data from the National Longitudinal Study of Youth 1979 (NLSY79). **Notes:** The distributions above represent cognitive ability factors estimated using a subset of the Armed Services Vocational Aptitude Battery (ASVAB) and educational attainment 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.

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 skills. On a variety of other dimensions, GED recipients behave much more like other dropouts. Figure 7 shows measures of early adolescent drug use, crime, sex, and violence extracted from three data sources.[69] 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. Figures 8a and b summarize 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.

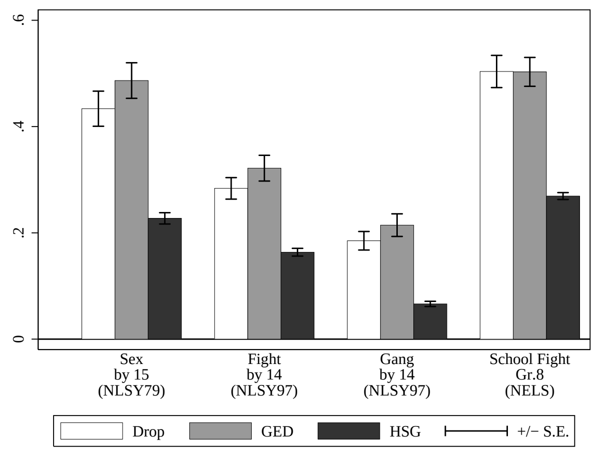
[69]See National Longitudinal Survey of Youth 1979 (NLSY79), National Longitudinal Survey of Youth 1997 (NLSY97), and National Educational Longitudinal Survey (NELS). For discussion of these data sets, see Heckman et al. (2014).

Figure 7 Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates

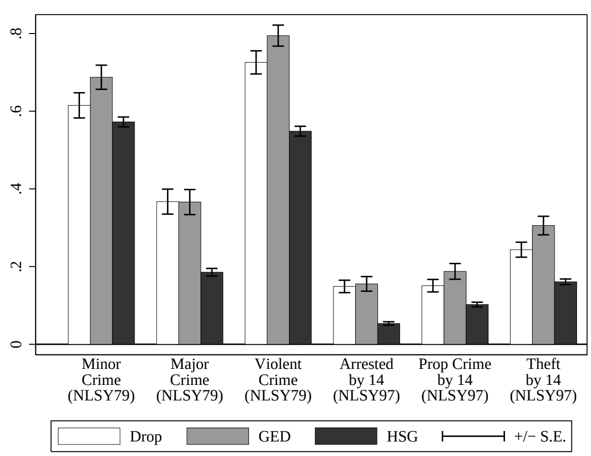
(a) Smoking and Drinking



(b) Sex and Violent Behavior



(c) Criminal Behavior

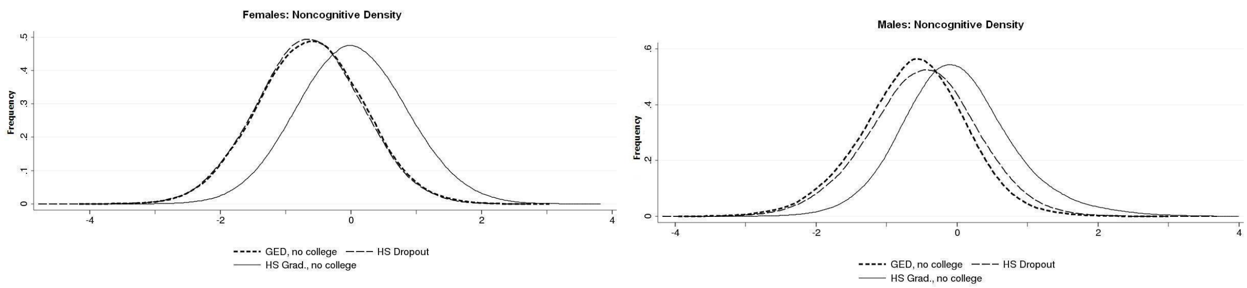


Sources: Heckman et al. (2014, Chapter 3). National Longitudinal Survey of Youth 1979, National Longitudinal Survey of Youth 1997, National Educational Longitudinal Survey. **Notes:** Minor crime includes vandalism, shoplifting, petty theft, fraud, holding or selling stolen goods. Major crime includes 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.

Figures 8a and b shows the distribution of a factor summarizing the diverse measures of adolescent risky behavior for dropouts, GED recipients, and high school graduates. On this index, GED recipients are nearly identical to high school dropouts.

Figure 8 Distribution of Non-Cognitive Skills by Education Group and Distribution of a Summary Measure of Noncognitive Ability by Education Group

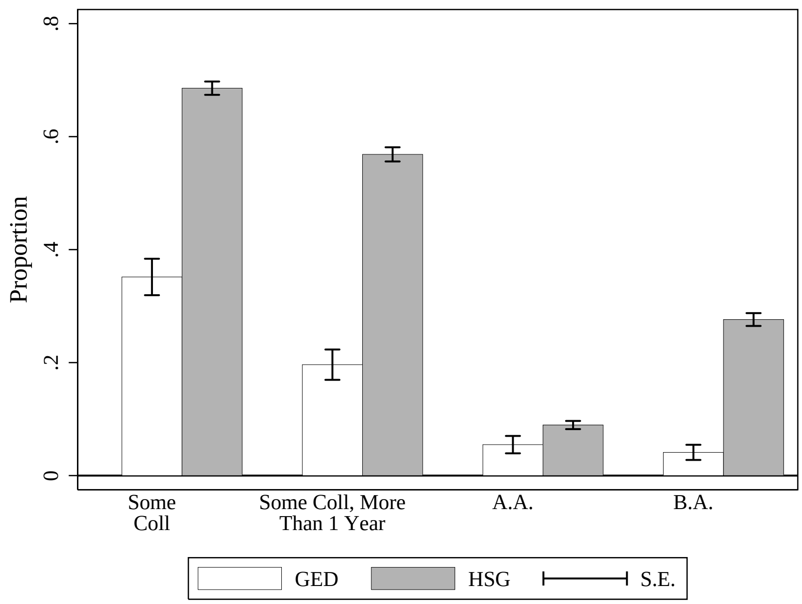
(a) Males **(b)** Females



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 skills 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 9 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 9 Post-Secondary Educational Attainment Across Education Groups Through Age 40 (NLSY79) - Males

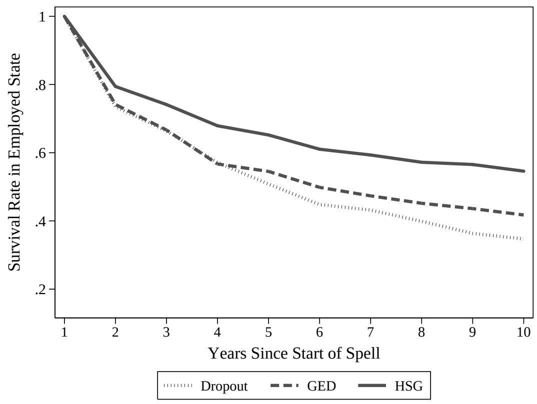


Sources: Heckman et al. (2014, Chapter 4). National Longitudinal Survey of Youth 1979. **Notes:** The graph shows post-secondary educational attainment of GED recipients and high school graduates. **Variable Definitions:** "Some College" represents people who entered any post-secondary institution ever. "Some College, More Than a Year" represents people who completed at least a year of some post-secondary education ever. "A.A." represents people who obtained associate's degrees ever. "B.A." represents people who obtained bachelor's degrees ever. "B.A." also includes people with higher education: M.A. Ph.D and professional degrees. **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.

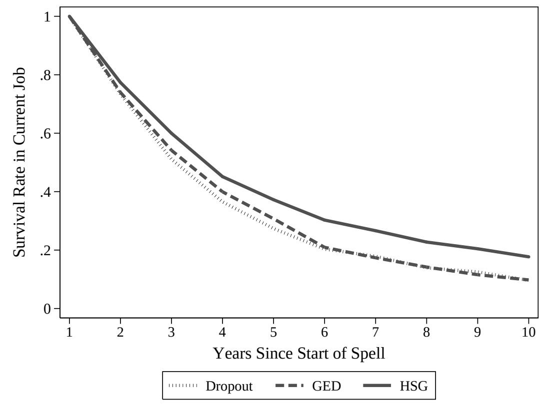
GED recipients lack persistence in a variety of tasks in life. Figure 10 shows the survival rates in employment (overall), employment in a given job, marriage, and in the condition of not having been incarcerated. GED recipients tend to exit employment, become divorced, and enter jail at rates similar to those of high school dropouts, while high school graduates are much more persistent.

Figure 10 Survival Rates in Various States for Male Dropouts, GED Recipients, and High School Graduates

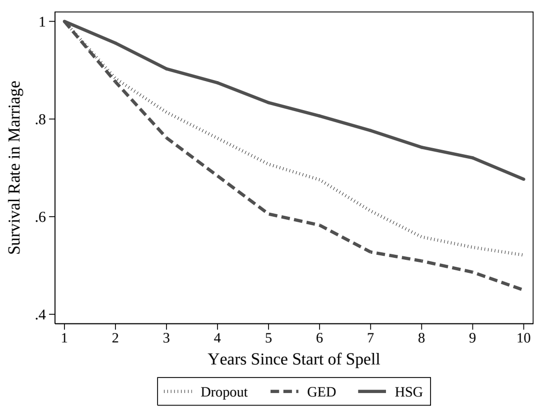
(a) Survival Rate in Employment



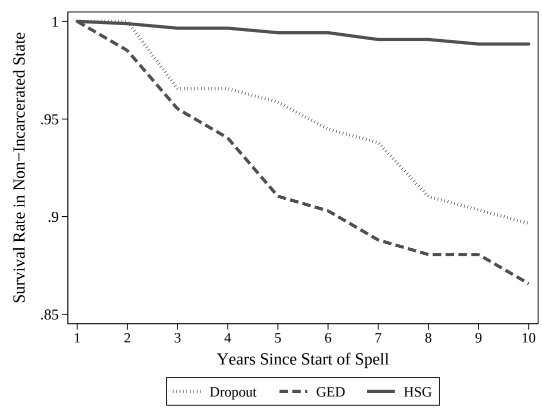
(b) Survival Rate in Staying on a Job



(c) Survival Rate in Marriage



(d) Survival Rate in Staying Out of Jail



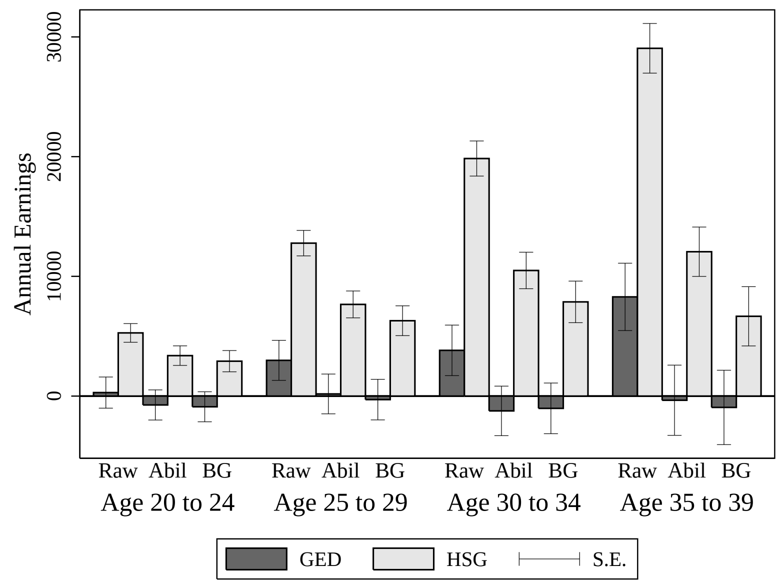
Source: Heckman et al. (2014, Chapter 4). National Longitudinal Survey of Youth 1979 (NLSY79), nationally representative cross sectional sample. Notes: The spell to first time being incarcerated begins in the first year that individuals exit school. **Tests of Significance:** The estimates for GED recipients 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 11 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. Adjusting for family background characteristics does not change the story.[70]

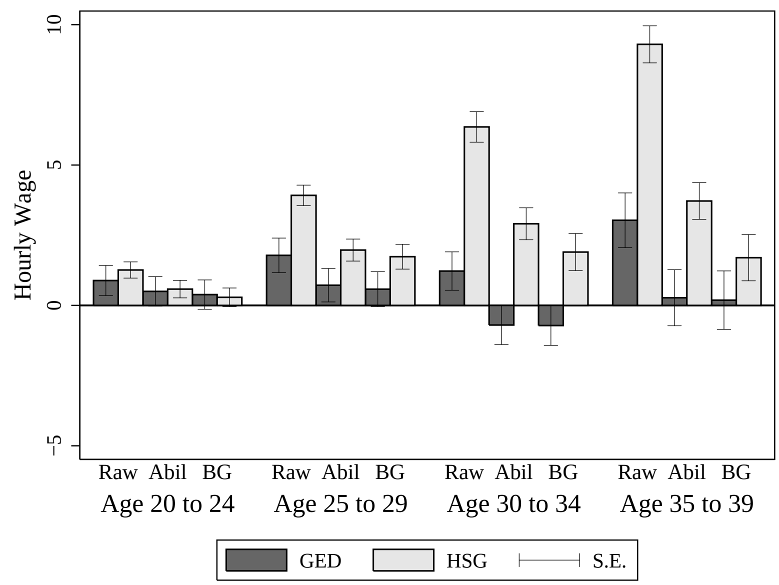
[70] Most of the patterns found for women parallel those found for men. However, there are some important differences.[71] While female GED recipients share similar cognitive and personality skills as male GED recipients, their outcomes differ. After accounting for differences in cognitive ability, female 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. 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. (2014) for a full discussion of the evidence on the performance of GED recipients.

Figure 11 Labor Market Outcomes Differences - By Age - NLSY79 - Males

(a) Annual Earnings



(b) Hourly Wage



Source: Heckman et al. (2014, 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.

3 Causal Evidence from Intervention Studies

Predictive studies do not establish causality. Most studies in personality psychology do not address the question of causality, i.e., whether measured skills *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 skills *cause* outcomes.

We introduce a simple framework to analyze the effect of skills on outcomes and how skills evolve over time.[72] 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 such as education and job training K_a, and the effort allocated to the task e_{T_a}:

$$\underbrace{T_a}_{\substack{\text{Outcome on a} \\ \text{task at age } a}} = \phi_a(\underbrace{C_a}_{\text{Cognition}}, \underbrace{P_a}_{\text{Personality}}, \underbrace{K_a}_{\substack{\text{Other} \\ \text{acquired} \\ \text{skills}}}, \underbrace{e_{T_a}}_{\substack{\text{Effort} \\ \text{devoted to} \\ \text{task}}}) \quad a = 1, \ldots, A. \tag{1}$$

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 Υ_a:

$$e_{T_a} = \psi_{T_a}(C_a, P_a, K_a, \underbrace{R_{T_a}}_{\substack{\text{Incentives} \\ \text{to perform} \\ \text{on task}}}, \underbrace{\Upsilon_a}_{\text{Preferences}}). \tag{2}$$

This representation distinguishes preferences from skills, although as previously noted, the two are likely closely related. Their relationship is an ongoing topic of research.

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

[72] This framework draws on Almlund et al. (2011).

e.g., Baumeister and Tierney, 2011).

Equations (1) and (2) formalize the difficulty in establishing causal relationships between outcomes and skills. Multiple skills and effort 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 skill the analyst is trying to measure.[73] This practice can lead to a substantial bias in inference about any particular skill.

In addition, most studies assume a linear or at least monotonic relationship between outcomes and skills. This practice is particularly problematic for measuring personality skills, where the effect of a skill on an outcome is not always linear or monotonic. Too much of a good thing can be bad. For example, extreme levels of skills are associated with psychopathologies. High levels of conscientiousness are associated with obsessive-compulsive disorder, which hinders task performance (Samuel and Widiger, 2008). Nonlinearities can also arise when skills and incentives interact, as in the analyses of Borghans et al. (2008) and Segal (2008) who show that people with different personality skills respond differently to incentives on tests.[74]

Skills evolve over time through investment and habituation.[75] Equation (3) shows that skills at age $a+1$ are age-dependent functions of cognitive ability, personality skills, other acquired skills, and investment I_a at age a. In this way, previous levels of skills and acquired skill affect current levels of skills and acquired skill. Equation (3) formalizes the notion that skills governing performance at a point in time are themselves the outcome of investment and habituation:

$$(C_{a+1}, P_{a+1}, K_{a+1}) = \eta_a(C_a \;, \;\; P_a \;, \;\; K_a \;, \;\; \underbrace{I_a}_{\substack{\text{Investment}\\ \text{and}\\ \text{experience}}}), \qquad a = 1, \ldots, A. \tag{3}$$

[73]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).

[74]Formally, this occurs when $\frac{\partial^2 \psi_{T_a}}{\partial P_a \partial R_{T_a}} \neq 0$.

[75]Todd and Wolpin (2006) find that differences in parental investment (home inputs) explain almost a third of the Black-White and Hispanic-White achievement gap. See also Cunha and Heckman (2008a).

In conjunction with resource constraints, a "deeper" set of preference parameters at age a may govern investment decisions and effort allocated to tasks. See Figure 12 for a schematic.

In addition, investment today increases the stock of future skills, which in turn increases the return to future investments. Economists call this phenomenon *dynamic complementarity.* This channel increases the returns to early investments because it makes future investments more productive. For this reason, Cunha et al. (2010) show that it is economically efficient to invest in the most disadvantaged young children because it raises their payoffs from future investments. Heckman and Mosso (2014) present a more complete discussion of static and dynamic complementarity and a formal proof of when early investment is more effective compared to later investment.

An important extension of this modelling approach is that performance on current tasks themselves can depend directly on performance of past tasks independently of a person's skills or effort. This embodies the idea of habituation that was discussed by Aristotle[76]: constant practice of moral behavior can make persons moral habitually. Formally, Equation (1) can be modified as:

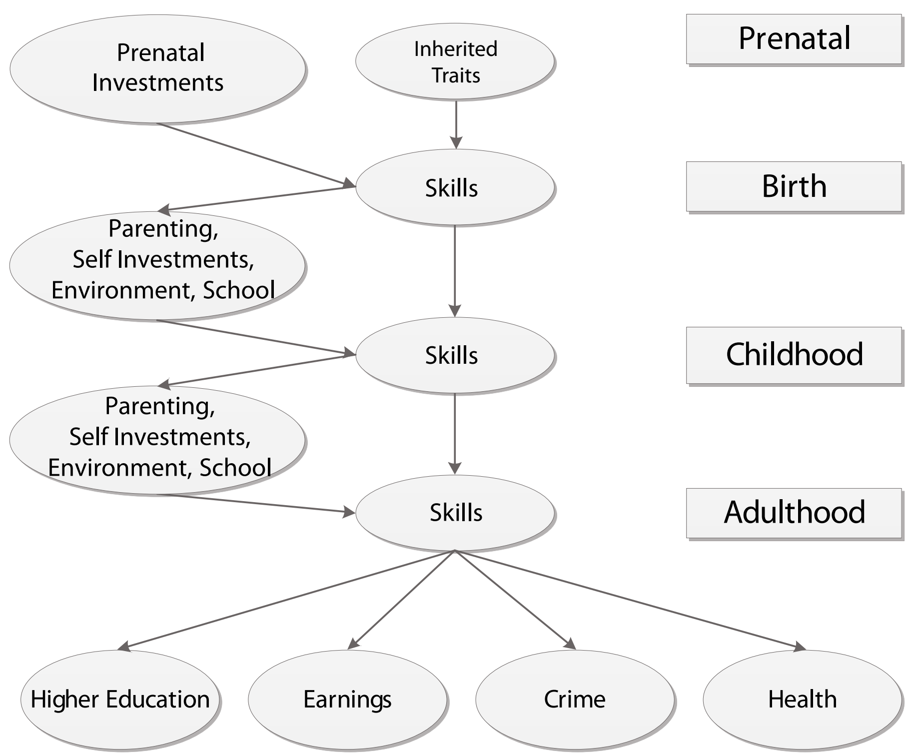
$$T_a = \phi_a^{'}(C_a, P_a, K_a, e_{T_a}, T_{a-1}^{'}) \tag{4}$$

where $T_{a-1}^{'}$ might represent a fundamentally different task that measures the same set of skills. For example, $T_{a-1}^{'}$ could represent the number of course credits that a student earns and T_a could represent graduating from high school. Both are valid measures of skills. As suggested by Jackson (2018) and Kautz and Zanoni (2019), credits earned is a valid measure of non-cognitive skills because successfully completing courses depends on non-cognitive skills. Similarly, as argued by Heckman et al. (2014), graduating from high school is a task that depends in part on non-cognitive skills (see d, Section III). A complication arises because students who earn more credits in 9th grade might be more likely to graduate from high school in part because they have higher levels of skills, but also because high school graduation is

[76]Ross (1956)

a direct function of credits earned. More generally, task T'_{a-1} captures something other than the underlying skills of students that can affect performance on T_a. Econometric methods can separate the direct of past measures from the effect of skills (Athans and Falb, 2013; Granger, 1969; Heckman, 1981a,c,b; Harvey, 1989; Torgovitsky, 2015; Williams, 2018). Models of this type are also common in the economics literature and are beneficial, because they capture learning (Becker and Murphy, 1988; Hai and Heckman, 2019; Pollak, 1976).

Figure 12 Framework for Understanding Skill Development



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

Figure 12 illustrates why understanding the effects attributable to specific interventions is such a challenging task. Most empirical studies only investigate the interventions aimed at one slice of the life cycle. They do not connect the links in the figure or correct for the effects of later investment in producing the outcomes attributed to early investments. One important area for future research on skill formation is to better document how early interventions influence the efficacy of later interventions.

3.1 Evidence from The Perry Preschool Program and Other Interventions

Evidence from the Perry Preschool Program shows how non-cognitive skills 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.[77]

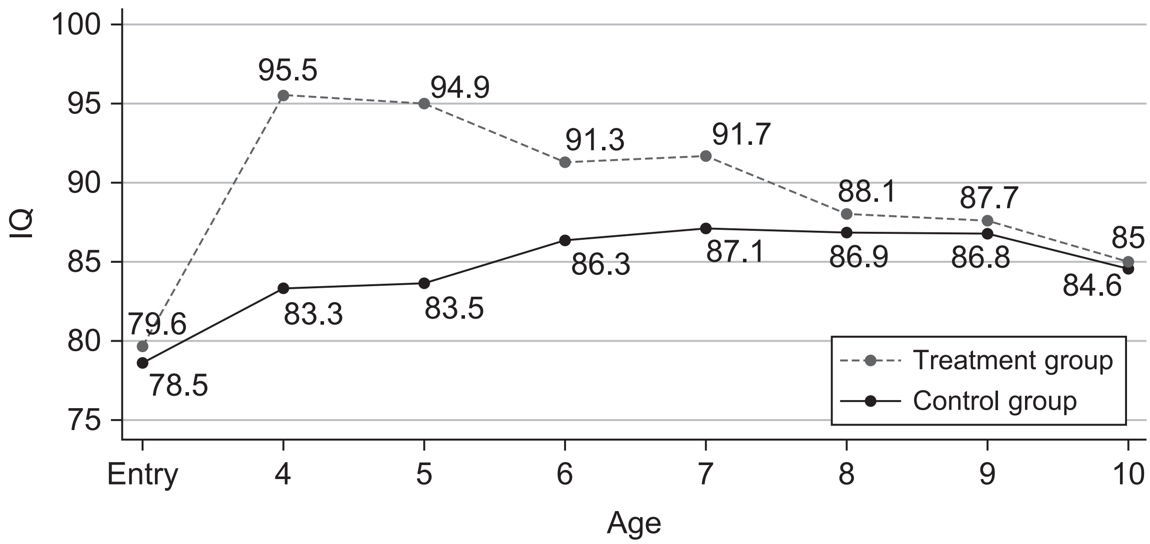
Participants were taught social skills in a "plan-do-review" sequence where students planned a task, executed it, and then reviewed it with teachers and fellow students. They learned to work with others when problems arose.[78] In addition, home visits promoted parent-child interactions. The program 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 13 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 intervention studies. Similar evidence from Head Start programs and a faith in IQ as a central determinant of life success strongly influenced Arthur Jensen's views about the genetic determination of skills (Jensen, 1969).

[77] We draw on the analysis of Heckman et al. (2008).

[78] Sylva (1997) describes the Perry program as a Vygotskian program fostering personality skills. Vygotsky developed a psychology of child development in structured social settings that emphasized development of social and personality skills.

Figure 13 Perry Preschool Program: IQ, by Age and Treatment Group



Notes: IQ measured on the Stanford-Binet Intelligence Scale (Terman and Merrill, 1960). The test was administered at program entry and at each of the ages indicated. Source: 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 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.[79]

The Perry Preschool Program worked primarily through improving personality skills. 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 supports the evidence previously presented that shows that performance on

[79]See DeLong and Magin (2009).

achievement tests depends in part on personality skills. Arthur Jensen's lifetime campaign against early intervention program was based on using faulty measures of relevant lifetime skills.

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.[80] 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).[81] 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,

[80]Bierman et al. (2010).

[81]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).

2006).[82] Positive findings are reported for the Montessori preschool curriculum (Lillard and Else-Quest, 2006). Unlike the Perry study, these studies do not have long-term followups.

There is evidence that targeted intervention efforts can improve preferences and skills. In contrast to the multi-faceted curricula described above, studies targeting improvement in aspects of conscientiousness are designed to isolate a particular mechanism producing behavioral change. In early work, 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). While yielding interesting preliminary results, these programs had only short-term follow-ups and involved very small samples.

A recent strand of research has focused on testing and implementing interventions on a larger scale and directly in school. Alan and Ertac (2018) show that an intervention in Turkey designed to encourage forward-looking behavior by increasing the salience of future selves improves patience measured on experimental tasks. The effect is persistent three years later and associated with an improved "behavior grade" for girls and high achieving students. Kosse et al. (2014) show that social skills can also be impacted. They study a mentoring intervention program in Germany which randomly paired children from low income families with college student mentors. Before the intervention, children from low income families scored lower on measures of prosociality. The gap across income groups disappeared for the treated group was still found two years later. Finally, Algan et al. (2014) study a two-year intervention (implemented at the time of entry into primary school) aimed at improving non-cognitive skills in boys who were disruptive in kindergarten. They find that it increased self-control and trust; improved labor market participation; and reduced criminal behavior

[82] However, a more recent large-scale study (Farran et al., 2011) does not find any effect of the program on self-regulation or literacy, language, and mathematics achievement.

in early adulthood.

Several studies suggest that non-cognitive skills 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. Kautz and Zanoni (2019) find similar effects for a mentoring program in high schools in disadvantagef neighborhoods.

Heckman et al. (2006) estimate a version of Equation (3) to analyze the effects of increases in education on measured cognition and non-cognitive measures.[83] Controlling for the problem of reverse causality that schooling may be caused by non-cognitive skills, they find that schooling improves both personality and cognitive skills and that these skills, in turn, boost outcomes.[84] Heckman et al. (2018) 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 skills affect schooling choices, labor market outcomes, adult health, and social outcomes and that increasing education promotes beneficial labor market, health, and social outcomes.

Todd and Zhang (2019) confirm that returns to schooling are in part a consequence of positive changes to personality through education. They find that these changes are concentrated predominantly among individuals from poorer families and tend to stabilize by the age of 30. Furthermore, some authors claim that cognitive and non-cognitive skills are associated with sorting into different job types and individuals who score high on both tend to choose more schooling and subsequent employment in white collar occupations.

[83] They estimate the effect of schooling on self-esteem and locus of control, personality skills related to neuroticism. The Rosenberg Self-Esteem Scale attempts to assess the degree of approval or disapproval of oneself (Rosenberg, 1965). The relationship between these measures and the Big Five skills of neuroticism is discussed in Almlund et al. (2011).

[84] Both Heckman et al. (2011) and Heckman et al. (2006) use an identification strategy based on matching on proxies for unobserved skills that corrects for measurement error and the endogeneity of schooling.

Kassenboehmer et al. (2018) contribute to this literature and provide estimates of the effect of university education on the Big Five personality skills. Controlling for selection into college, they show that it increases the extraversion skill by 0.3 standard deviations and seems to have some impact on agreeableness although the later is quite heterogeneous and depends on family background.

Cunha et al. (2010) 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 skills. They control for the endogeneity of investment using shocks to family income along with other instruments. Their model is a version of Equation (3). Skills are self-productive and exhibit dynamic complementarity – current values of skills affect the evolution of future skills through direct and cross effects. A leading example of a cross effect is that more motivated children are more likely to learn.[85] They estimate parameters that summarize how past personality skills affect future cognitive skills.

They find that self-productivity becomes stronger as children become older, for both cognitive and personality skills.[86] 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 Cunha et al. (2006); Knudsen et al. (2006) and Cunha and Heckman (2007).

Personality skills foster the development of cognition but not vice versa (see Cunha and Heckman, 2008a and Cunha et al., 2010). It is relatively easier at all stages of life to

[85]There is preliminary evidence that the personality of one's peers may also have an impact on his outcomes. Golsteyn et al. (2017) exploit random variation in assignment of students to university tutorial sections to estimate a positive effect on performance from having more persistent and more risk averse peers. This effect is limited to students who themselves have a low degree of persistency and is twice as large in magnitude than that of having peers with higher GPA. As the hours spent studying are unaffected, the authors conclude that the presence of persistent and risk averse peers directly enhances the productivity of low-persistence students in their company.

[86]In the language of economics, the elasticity of substitution for cognitive inputs is *smaller* later in life.

compensate for early disadvantage in endowments by boosting personality skills.[87] However, personality seems to stabilize around the age of 30.[88] Thus, the most effective adolescent interventions target personality skills.[89]

Some life experiences, like employment, can also 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).[90] 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 than those in the control group. After 36 months, those who received the subsidy were more likely to have an improved locus of control.

Negative life experiences can have lasting effects on preferences and personality as well. Americans who experienced sexual abuse and parental neglect in childhood appear to have increased levels of neuroticism and lower conscientiousness and openness to experience at age 30 (Fletcher and Schurer, 2017). Afghanis who experienced violence exhibit more risk tolerance but also a higher preference for certainty when asked to recall fearful events (Callen et al., 2014). Furthermore, Malmendier and Nagel (Malmendier and Nagel) document that individuals who experienced negative financial events such as the Great Depression exhibit a lower willingness to take financial risks.[91] Anger et al. (2017) show that trauma can sometimes also result in positive changes in personality. Studying German data, they find

[87] Elasticities of substitution are essentially the same at different stages of the life cycle.

[88] See Todd and Zhang (2019) and Terracciano et al. (2010).

[89] Cunha et al. (2006) report that 16% of the variation in educational attainment is explained by adolescent cognitive skills, 12% is due to adolescent personality (socioemotional skills), and 15% is due to measured parental investments.

[90] The relationship between locus of control and the Big Five trait of neuroticism is discussed in Almlund et al. (2011).

[91] However, it is unclear what part of these changes can be attributed to changes in risk preferences as opposed to altered beliefs about returns to investing.

that job loss due to factory closings increases openness to experience for highly educated workers while leaving other dimensions of personality unchanged.

Economic preferences have also been shown to have a causal impact on outcomes.[92] Montizaan et al. (2015) exploit a change in the Dutch public pension system in 2006 which affected workers born after 1950. By comparing the reaction of public sector workers born just after the reform took effect to those born just before, they are able to show that affected individuals who score higher on negative reciprocity reduced their work effort (measured by self-reported on the job motivation). Furthermore, this decline seems proportional to the degree of perceived unfairness – it is larger for workers born very close to the cutoff date and among those who work with many colleagues who were unaffected – and also to the closeness to the "perpetrator" of the injustice (workers working directly for the central government shirk more).

4 Summary

This paper reviews recent evidence from economics and psychology on the importance of personality. It shows that success in life depends on many skills, not just those measured by IQ, grades, and standardized achievements tests. Personality skills predict and *cause* outcomes.

Economic analysis clarifies psychological studies by establishing that psychological attributes are measured by performance on tasks. It enriches the analysis of human differences by providing anchored measures of economic preferences and studying their links to personality and cognitive skills. Psychological attributes have different productivities in different tasks. Performance on tasks depends on incentives and multiple skills, giving rise to a fundamental identification problem when measuring any single skill. This identification problem is empirically important even for measures of cognitive skills.

[92] As discussed before, recent research suggests that they may be strongly related to non-cognitive personality skills traditionally measured by psychologists.

The importance of cognitive and non-cognitive skills increases with the complexity of a task. Given their endowments of skills and the incentives they face, people sort into tasks in life in pursuit of their comparative advantage. Recent research by Caines et al. (2017a) and Beaudry et al. (2016) shows that labor market rewards are increasing for those who have large *bundles* of both social and cognitive skills. They perform complex tasks. The demand for such bundles is increasing, not the demand for cognitive or non-cognitive skills alone, despite some recent claims to the contrary. In fact, there is evidence that the demand for cognitive skills alone is in decline (see Caines et al., 2017a,b, and Beaudry et al., 2016), and the evidence for a rise in the demand for non-cognitive skills by themselves is debated in the literature.

Skills are stable across situations. However, their manifestations depend on incentives to apply effort in the situations where they are measured and also on other skills.

However, skills are not set in stone. They change over the life cycle and can be enhanced by education, mentoring, parenting, and environmental influences to different degrees at different ages.

Scores on achievement tests capture both cognitive and personality skills. 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 studies surveyed here demonstrate that the growing partnership between economists and psychologists is contributing greatly to knowledge. Each field has a lot to learn from each other. Recent evidence changes the way we think about the sources of individual differences in life performance and ways to design effective policies for equalizing opportunities.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for truth-telling. *Econometrica 87*(4), 1115–1153.

Ackerman, P. L. and E. D. Heggestad (1997, March). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin 121*(2), 219–245.

Alan, S. and S. Ertac (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy 126*(5), 1865–1911.

Algan, Y., E. Beasley, F. Vitaro, and R. E. Tremblay (2014). The impact of non-cognitive skills training on academic and non-academic trajectories: From childhood to early adulthood. Unpublished.

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

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences. *Econometrica 76*(3), 583–618.

Anderson, J., S. Burks, C. DeYoung, and A. Rustichini (2011). Toward the integration of personality theory and decision theory in the explanation of economic behavior. Unpublished manuscript. Presented at the IZA Workshop: Cognitive and Non-Cognitive Skills, January 27, 2011.

Andersson, O., H. J. Holm, J.-R. Tyran, and E. Wengstr´om (2018). Robust inference in risk elicitation tasks. Unpublished.

Anger, S., G. Camehl, and F. Peter (2017). Involuntary job loss and changes in personality traits. *Journal of Economic Psychology 60*, 71–91.

Antecol, H. and D. A.Cobb-Clark (2013). Do psychosocial traits help explain gender segregation in young people's occupations? *Labour Economics 21*, 59–73.

Athans, M. and P. L. Falb (2013). *Optimal control: an introduction to the theory and its applications*. Courier Corporation.

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics 130*(3), 1067–1115.

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

Baldwin, J. (1995). *Who Took the GED? GED 1994 Statistical Report*. Washington, DC: American Council on Education, GED Testing Service, Center for Adult Learning.

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

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

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

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

Baumeister, R. F., A. L. Boone, and J. P. Tangney (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality 72*(2), 271–324.

Baumeister, R. F. and J. J. M. Tierney (2011). *Willpower : rediscovering the greatest human strength*. New York: Penguin Press.

Beauchamp, J. P., D. Cesarini, and M. Johannesson (2017). The psychometric and empirical properties of measures of risk preferences. *Journal of Risk and Uncertainty 54*(3), 203–237.

Beaudry, P., D. A. Green, and B. M. Sand (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics 34*(S1), S199–S247.

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

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (1 ed.). Chicago: University of Chicago Press for the National Bureau of Economic Research.

Becker, G. S. and K. M. Murphy (1988, August). A theory of rational addiction. *Journal of Political Economy 96*(4), 675–700.

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

Bierman, K. L., J. D. Coie, K. A. Dodge, M. T. Greenberg, J. E. Lochman, R. J. McMahon, and E. Pinderhughes (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology 78*(2), 156–168.

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

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

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

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

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2011b). IQ, achievement, and personality. Unpublished manuscript, University of Maastricht and University of Chicago (revised from the 2009 version).

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

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

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

Bowen, W. G., M. M. Chingos, and M. S. McPherson (2009). Test scores and high school grades as predictors. In *Crossing the Finish Line: Completing College at America's Public Universities*, Chapter 6, pp. 112–133. Princeton, NJ: Princeton University Press.

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

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

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

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

Cadena, B. C. and B. J. Keys (2015). Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy 7*(3), 126–53.

Caines, C., F. Hoffmann, and G. Kambourov (2017a). Complex-task biased technological change and the labor market. *Review of Economic Dynamics 25*, 298–319.

Caines, C., F. Hoffmann, and G. Kambourov (2017b). Training programs, skills, and human capital: A life-cycle approach. Unpublished.

Caliendo, M., D. A. Cobb-Clark, and A. Uhlendorff (2015). Locus of control and job search strategies. *Review of Economics and Statistics 97*(1), 88–103.

Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and risk preference: Experimental evidence from Afghanistan. *American Economic Review 104*(1), 123–48.

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

Castillo, M., P. J. Ferraro, J. L. Jordan, and R. Petrie (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics 95*(11–12), 1377–1385.

Castillo, M., J. L. Jordan, and R. Petrie (2018). Childrens rationality, risk attitudes and field behavior. *European Economic Review 102*, 62–81.

Chapman, J., M. Dean, P. Ortoleva, E. Snowberg, and C. Camerer (2018). Econographics. NBER Working Paper No. 24931.

Chen, Y., S. Feng, J. J. Heckman, and T. Kautz (2019). Sensitivity of self-reported non-cognitive skills to survey administration conditions. Forthcoming, *Proceedings of the National Academy of Sciences*.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review 105*(5), 1–33.

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

Choi, S., S. Kariv, W. Müller, and D. Silverman (2014). Who is (more) rational? *American Economic Review 104*(6), 1518–1550.

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

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

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamh´ofer, and H. Schildberg-H´orisch (2019). Self-control: Determinants, life outcomes and intergenerational implications. DICE Discussion Paper 319, D´usseldorf Institute for Competition Economics (DICE).

Cobb-Clark, D. A., S. C. Kassenboehmer, and M. G. Sinning (2016). Locus of control and savings. *Journal of Banking & Finance 73*, 113–130.

Cobb-Clark, D. A. and M. Tan (2011, January). Noncognitive skills, occupational attainment, and relative wages. *Labour Economics 18*(1), 1–13.

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

Credé, M., M. C. Tynan, and P. D. Harms (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology 113*(3), 492–511.

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

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

Cunha, F. and J. J. Heckman (2008b, September). A new framework for the analysis of inequality. *Macroeconomic Dynamics 12*(2 (supplement)), 315–354.

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

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

Daly, M., C. P. Harmon, and L. Delaney (2009, April–May). Psychological and biological foundations of time preference. *Journal of the European Economic Association 7*(2–3), 659–669.

Dean, M. and P. Ortoleva (2019). The empirical relationship between nonstandard economic behaviors. *Proceedings of the National Academy of Sciences 116*(33), 16262–16267.

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

Dee, T. S. and M. R. West (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis 33*(1), 23–46.

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

Deming, D. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

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

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010, June). Are risk aversion and impatience related to cognitive ability? *American Economic Review 100*(3), 1238–1260.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives 32*(2), 115–34.

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

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

Duckworth, A. L. and P. D. Quinn (2009, March). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment 91*(2), 166–174.

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

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

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

El-Gamal, M. A. and D. M. Grether (2012). Are people Bayesian? uncovering behavioral strategies. *Journal of the American Statistical Association 90*(432), 1137–1145.

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

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA Discussion Paper No. 9674.

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

Fehr, E. and L. Goette (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review 97*(1), 298–317.

Fletcher, J. M. and S. Schurer (2017). Origins of adulthood personality: The role of adverse childhood experiences. *The B.E. Journal of Economic Analysis and Policy 17*(2).

Flinn, C., P. Todd, and W. Zhang (2019). Personality traits, job search and the gender wage gap. Unpublished.

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

Forsythe, R., J. Horowitz, N. L.Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior 6*(3), 347–369.

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

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

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics 129*(4), 1661–1710.

García, E. (2014). The need to address noncognitive skills in the education policy agenda. EPI Briefing Paper 386, Economic Policy Institute, Washington, D. C.

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

Golsteyn, B., A. Non, and U. Zölitz (2017). The impact of peer personality on academic achievement. *University of Zurich, Department of Economics, Working Paper* (269).

Golsteyn, B. and H. Schildberg-Hörisch (2017). *Challenges in Research on Preferences and Personality Traits: Measurement, Stability, and Inference*. Elsevier.

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014). Adolescent time preferences predict lifetime outcomes. *The Economic Journal 124*(580), F739–F761.

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

Granger, C. W. J. (1969, August). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 424–438.

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

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

Hai, R. and J. J. Heckman (2019). A dynamic model of health, education, and wealth with credit constraints and rational addiction. University of Chicago.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004, July–August). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1–2), 39–98.

Harrison, G. W. (2006). Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics 34*(1), 125–162.

Harrison, G. W. and E. Elisabet Rutström (2008). Risk aversion in the laboratory. In *Risk Aversion in Experiments*, pp. 41–196. Emerald Group Publishing Limited.

Harrison, G. W. and J. T. Swarthout (2016). Cumulative prospect theory in the laboratory: A reconsideration. *Center for the Economic Analysis of Risk, Working Paper 2*.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.

Heckman, J. J. (1981a). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in Labor Markets*, National Bureau of Economic Research, pp. 91–139. University of Chicago Press.

Heckman, J. J. (1981b). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. In C.Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data With Econometric Applications*, pp. 179–195. Cambridge, MA: MIT Press.

Heckman, J. J. (1981c). Statistical models for discrete panel data. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, MA: MIT Press.

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

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

Heckman, J. J., J. E. Humphries, and G. Veramendi (2014). Education, health and wages. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health and smoking. *Journal of Political Economy 126*(S1), S197–S246.

Heckman, J. J. and T. Kautz (2012, August). Hard evidence on soft skills. *Labour Economics 19*(4: European Association of Labour Economists 23rd Annual Conference, Paphos, Cyprus, 22–24th September 2011), 451–464.

Heckman, J. J. and T. Kautz (2014a). Achievement tests and the role of character in American life. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, Chapter 1, pp. 3–56. Chicago: University of Chicago Press.

Heckman, J. J. and T. Kautz (2014b). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, Chapter 9, pp. 341–430. Chicago: University of Chicago Press.

Heckman, J. J., L. Malofeeva, R. Pinto, and P. A. Savelyev (2008). The effect of the Perry Preschool Program on cognitive and noncognitive skills: Beyond treatment effects. Unpublished manuscript, University of Chicago, Department of Economics.

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

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

Heckman, J. J. and S. Mosso (2014, August). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

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

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

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

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

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

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

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

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

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

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

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jagelka, T. (2018). Are economists' preferences psychologists' personality traits? Working Paper.

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

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review 39*(1), 1–123.

John, O. (2000). The big five personality test. `http://www.outofservice.com/bigfive/?score-bigfive`.

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

John, O. P. and S. Srivastava (1999a). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2 ed.)., Chapter 4, pp. 102–138. New York: The Guilford Press.

John, O. P. and S. Srivastava (1999b). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research 2*(1999), 102–138.

Jones, M. A. (2018). The middle-skills gap: Community-based solutions to meet employer demands. Technical Report Chapter 12 in *Investing in America's Workforce: Improving Outcomes for Workers and Employers*, Atlanta Federal Reserve.

Kassenboehmer, S. C., F. Leung, and S. Schurer (2018). University education and non-cognitive skill development. *Oxford Economic Papers 70*(2), 538–562.

Kautz, T. and W. Zanoni (2019). Measuring and fostering non-cognitive skills in adolescence: Evidence from Chicago Public Schools and the OneGoal Program. Unpublished manuscript, University of Chicago, Department of Economics.

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

Kettner, N. (1976). Armed Services Vocational Aptitude Battery (ASVAB form 5): Comparison with GATB and DAT tests: Final report for period May 1975–October 1976. Technical report, Department of Defense, Department of the Air Force, Air Force Systems Command, Air Force Human Resources Laboratory, Brooks Air Force Base, TX. Prepared by Contemporary Research, Incorporated with Personnel Research Division, Air Force Human Resources Laboratory, Lackland Air Force Base for Air Force Human Resources Laboratory, Brooks Air Force Base.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

Kosse, F., T. Deckers, H. Schildberg-Hörisch, and A. Falk (2014). Formation of human prosociality: Causal evidence on the role of social environment. University of Bonn, Germany.

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

Kyllonen, P. C. and J. P. Bertling (2013). Innovative questionnaire assessment methods to increase cross-country comparability. *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 277–285.

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

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

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

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

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

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

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and E. S. Welch (2001). Risk as feelings. *Psychological Bulletin 127*, 267–286.

Malmendier, U. and S. Nagel. Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

Mandelbrot, B. B. (1962). Paretian distributions and income maximization. *Quarterly Journal of Economics 76*, 57–85.

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

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

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

Means, B. and J. H. Laurence (1984, May). Characteristics and performance of recruits enlisted with General Educational Development (GED) credentials. Final Report 84-6, Human Resources Research Organization, Alexandria, VA.

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

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

Montizaan, R., A. de Grip, F. Cörvers, and T. Dohmen (2015). The impact of negatively reciprocal inclinations on worker behavior: Evidence from a retrenchment of pension rights. *Management Science 62*(3), 668–681.

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

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

Nelsen, B. (1997). Should social skills be in the vocational curriculum? Evidence from the automotive repair field. In A. M. Lesgold, M. J. Feuer, and A. M. Black (Eds.), *Transitions in Work and Learning: Implications for Assessment*, Chapter 4, pp. 62–88. Washington, DC: National Academy Press.

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

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

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

OECD (2013a). PISA 2012 results: Ready to learn: Students engagement, drive and self-beliefs (volume iii). Technical report, OECD Publishing.

OECD (2013b). What is PISA? In *PISA 2012 Results: What Students Know and Can Do*, Volume I of *Student Performance in Mathematics, Reading and Science*, Chapter 1, pp. 23–30. OECD Publishing.

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

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

Pollak, R. A. (1976). Interdependent preferences. *The American Economic Review*, 309–320.

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

Primi, R., C. Zanon, D. Santos, F. De Fruyt, and O. P. John (2016). Anchoring vignettes. *European Journal of Psychological Assessment*.

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

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

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

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

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

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

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

Ross, D. (1956). Aristotle: The Nicomachean ethics. *Philosophy 31*(116), 77–77.

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

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied 80*(1), 1–28.

Roy, A. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rueda, M. R., M. K. Rothbart, B. D. McCandliss, L. Saccomanno, and M. I. Posner (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences 102*(41), 14931–14936.

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

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

Samuel, D. B. and T. A. Widiger (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review 28*(8), 1326–1342.

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

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

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

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

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

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

Segal, C. (2012, August). Working when no one is watching: Motivation, test scores, and economic success. *Management Science 58*(8), 1438–1457.

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

Spence, A. M. (1972). *Market Signalling: The Informational Structure of Job Markets and Related Phenomena*. Ph. D. thesis, Harvard University, Cambridge, MA.

Stango, V. and J. Zinman (2019). We are all behavioral, more or less: Measuring and using consumer-level behavioral sufficient statistics. NBER Working Paper No. 25540.

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

Stevens, C., J. Fanning, D. Coch, L. Sanders, and H. Neville (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Research 1205*, 55–69.

Sutter, M., C. Zoller, and D. Glätzle-Rützler (2019). Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review 111*, 98–121.

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

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

Terracciano, A., R. R. McCrae, and P. T. Costa Jr. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality 44*(1), 31–37.

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

Thurstone, L. L. (1927, July). A law of comparative judgement. *Psychological Review 34*(4), 273–286.

Todd, P. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in Mexico using experimental data to validate a dynamic behavioral model of child schooling. *American Economic Review 96*(5), 1384–1417.

Todd, P. and W. Zhang (2019). A dynamic model of personality, schooling, and occupational choice. Forthcoming, *Quantitative Economics.*

Torgovitsky, A. (2015, May). Identification of nonseparable models using instruments with small support. *Econometrica 83*(3), 1185–1197.

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

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

West, M. R., K. Buckley, S. B. Krachman, and N. Bookman (2018). Development and implementation of student social-emotional surveys in the core districts. *Journal of Applied Developmental Psychology 55*, 119–129.

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

Williams, B. (2018). Identification of the linear factor model. George Washington University, Department of Economics. Unpublished manuscript.

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