HCEO WORKING PAPER SERIES

What Grades and Achievement Tests Measure

Lex Borghans
Bart H. H. Golsteyn
James J. Heckman
John Eric Humphries

Working Paper

2016-022

11/2016

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

What Grades and Achievement Tests Measure*

Lex Borghans
Maastricht University

Bart H.H. Golsteyn
Maastricht University

James J. Heckman
University of Chicago
American Bar Foundation

John Eric Humphries
University of Chicago

November 10, 2016

*The names of the authors are in alphabetical order. This research was supported by: the American Bar Foundation; the Pritzker Children's Initiative; the Buffett Early Childhood Fund; NIH grants NICHD R37HD065072, NICHD R01HD54702, and NIA R24AG048081; an anonymous funder; Successful Pathways from School to Work, an initiative of the University of Chicago's Committee on Education; the Hymen Milgrom Supporting Organization; the Human Capital and Economic Opportunity Global Working Group, an initiative of the Center for the Economics of Human Development; the Institute for New Economic Thinking; a Tore Browaldh grant from the Handelsbanken Research foundation; and a VIDI grant from the Netherlands Organization for Scientific Research. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The authors received valuable comments from two referees, Anja Achtziger, Thomas Dohmen, Angela Lee Duckworth, Brent Roberts, Wendy Johnson, Tim Kautz, Anna Sjögren, seminar participants at IFAU (2010), SOFI (2010), IZA (2009), IFS (2009), Geary Institute (2009), Tilburg University (2009), the University of Konstanz (2009), and conference participants at the 2011 American Economic Association in Denver, the 2011 IZA conference on cognitive and noncognitive skills in Bonn, three Spencer conferences at the University of Chicago (2009 and 2010), the Second Conference on Non-Cognitive Skills in Konstanz, Germany (2009), the 2009 European Summer Symposium in Labour Economics in Buch am Ammersee, Germany, the 2009 meeting of the Association for Research in Personality in Evanston, IL, and the 2010 Society of Labor Economists/European Association of Labour Economists in London. A web appendix for this paper can be found at https://heckman.uchicago.edu/what-do-grades-measure. *Lex Borghans*: Department of Economics and Research Centre for Education and the Labour Market (ROA), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands, lex.borghans@maastrichtuniversity.nl. *Bart H. H. Golsteyn*: Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands, b.golsteyn@maastrichtuniversity.nl. *James J. Heckman*: Department of Economics, The University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, USA, jjh@uchicago.edu. *John Eric Humphries*: Department of Economics, The University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, USA, johneric@uchicago.edu.

Abstract

Intelligence quotient (IQ), grades, and scores on achievement tests are widely used as measures of cognition, yet the correlations among them are far from perfect. This paper uses a variety of data sets to show that personality and IQ predict grades and scores on achievement tests. Personality is relatively more important in predicting grades than scores on achievement tests. IQ is relatively more important in predicting scores on achievement tests. Personality is generally more predictive than IQ of a variety of important life outcomes. Both grades and achievement tests are substantially better predictors of important life outcomes than IQ. The reason is that both capture personality traits that have independent predictive power beyond that of IQ.

Keywords: : IQ, Achievement tests, Grades, Personality traits

JEL codes: J24, D03

1 Introduction

Intelligence quotient (IQ), grades, and scores on achievement tests are widely used as measures of cognition.[1] Yet, the correlations among them are far from perfect. This paper establishes the predictive power of personality for grades and scores on achievement tests. Personality is a better predictor of a variety of life outcomes than IQ. Both grades and scores on achievement tests have independent predictive power above and beyond IQ because both measures capture aspects of personality.

Achievement tests were designed to capture general knowledge acquired in school and life (see, e.g., Lindquist, 1951; Heckman and Kautz, 2012, 2014). They were thought to be more objective and fairer than grades, which involve teacher assessments of individual students in particular classrooms. Tests of fluid intelligence were designed to capture "innate aptitudes," rather than acquired knowledge (Green, 1974).

The recent literature has shown that there is no clear distinction between innate and acquired traits. A large body of research shows that IQ can be altered by interventions (see Almlund et al., 2011 and Elango et al., 2016). Additionally, all measures of ability are based on knowledge as gauged by performance on tasks (e.g., taking a test).[2] Not only is knowledge acquired, but greater cognitive ability facilitates acquisition of knowledge. Personality traits also affect acquisition of knowledge. More motivated people learn more (Borghans et al., 2008). In addition, more conscientious people take tests more seriously (Borghans et al., 2008). Personality traits also influence grades. It was precisely because grades depend on personality that achievement tests were advocated as better measures of cognition. Achievement tests were thought to be independent of teacher assessments of non-cognitive traits that were often deemed to be biased (Heckman and Kautz, 2012, 2014).

This paper makes the following points. (1) Grades, scores on achievement tests and IQ are strongly positively correlated, but not perfectly so. This strong correlation gives purchase

[1] See, e.g., Nisbett (2009) and Nisbett et al. (2012). Web Appendix 1 documents the widespread use of achievement tests as measures of IQ.

[2] See, e.g., Anastasi and Urbina (1997).

to the view that the three measures can be used interchangeably. (2) Grades and scores on achievement tests are differentially influenced by IQ and personality. Grades are more heavily influenced by personality than achievement tests. (3) All three measures predict a variety of important life outcomes, but scores on achievement tests and grades are better predictors than IQ. (4) Grades and achievement tests are more predictive of life outcomes because they capture aspects of personality that have independent predictive power.

The paper proceeds as follows: The first section briefly reviews the literature. The second section describes the data. The third section decomposes grades and scores on achievement tests into IQ and personality. The fourth section examines the predictive power of IQ and personality on a variety of important life outcomes.[3]

2 A brief overview of the literature

Achievement tests like the Armed Forces Qualification Test (AFQT) are often used as proxies for cognitive ability (see, e.g., Herrnstein and Murray, 1994, Murnane et al., 1995, and Hanushek and Woessmann, 2008). Web Appendix 1 lists 50 papers that use AFQT scores as proxies for intelligence. Grades are also used as proxies for intelligence (e.g., Nisbett, 2009 and Nisbett et al., 2012).

Previous research studies relationships between IQ and personality,[4] between grades and IQ,[5] and between personality and grades.[6] Barton et al. (1972) relate the High School

[3]We make no causal claims in this paper.

[4]Duckworth et al. (2011) give an overview of this literature. Scores on IQ tests have been related to personality (Borghans et al., 2009). In related work, Segal (2012) shows that less conscientious men perform better when they are offered incentives in IQ tests and Borghans et al. (2008) show that conscientious and emotionally stable people do not spend more time answering IQ questions when rewards are higher, while people who score lower on these traits do.

[5]Ackerman and Heggestad (1997) review the literature.

[6]Poropat (2009) and Poropat (2014) give an overview of this literature. Poropat (2009) concludes that Conscientiousness is the greatest Big Five predictor of grades (followed at some distance by Openness to Experience). Conscientiousness predicts academic performance almost as well as intelligence. Poropat (2014) evaluates how adolescent measures of the Big Five predict academic performance—finding that Openness and Conscientiousness are particularly important. Noftle and Robins (2007) investigate the relationship between verbal and mathematical Scholastic Aptitude Test (SAT) scores and the Big Five. They find that Openness to Experience relates to SAT verbal scores. See Almlund et al. (2011) for an extensive review.

Personality Questionnaire and the Culture Fair Intelligence Test to scores on standardized achievement tests and find that Conscientiousness and IQ predict scores on achievement tests. Duckworth and Carlson (2013) survey studies relating self-regulation and scores on standardized achievement tests, course grades, and high school achievement. They show that self-regulation is more predictive of course grades than scores on standardized achievement tests, and suggest that this may be the reason why course grades are more predictive of certain later-life outcomes than achievement tests. Duckworth and Seligman (2005) report that both self-discipline and IQ predict performance on achievement tests. Duckworth et al. (2012) report that self-control (a facet of Big Five Conscientiousness) and IQ (measured by Raven Matrices) predict scores on the English/language arts and mathematics standardized achievement tests. Our analysis builds on and extends this research by analyzing the effects of cognition and personality on grades, achievement tests, and a variety of important life outcomes. We report results from samples pooled across genders.

3 Data

Table 1 summarizes the availability of measures in the four data sets we analyze.[7] Although details and point estimates vary, and some data contain only partial information, consistent patterns emerge across all four data sets.

Stella Maris is a Dutch high school at which we collected Raven's IQ, scores on achievement tests (the Differential Aptitude Test, DAT), grades and measures of personality. For this sample, we have no measure of adult outcomes. The British Cohort Study (BCS) follows a cohort of children born in one week in April, 1970 until 2016. It has information on grades, IQ, scores on achievement tests, personality, and a variety of adult life outcomes. The NLSY79 samples American children aged 14–21 in 1979 and follows them ever since. It has an achievement test (the Armed Forces Qualifying Test, AFQT) and scores on different IQ tests across students, which we equate to produce a common IQ score. It has limited measures of

[7] Across data sets, the survey instruments differ somewhat. The definitions are given in the Web Appendix.

personality, but rich data on adult outcomes. The National Survey of Midlife Development in the United States (MIDUS) is a survey of adults aged 24–74 in 1995/6 and 34–83 in 2004/6. It has rich data on IQ, personality, and adult outcomes, but lacks information on achievement scores or grades. No single data set produces definitive evidence. It is the consilience of the evidence across the diverse data sets that justifies the conclusions of this paper.[8]

4 Grades, achievement tests, and personality

This section summarizes the correlations among the dimensions of human capabilities that we study. It also analyzes the extent to which personality predicts achievement test scores and grades above and beyond IQ.

Table 2 displays the correlations among the available measures of cognition and personality in our four data sets. Notice that the correlations between IQ and grades, as well as IQ and achievement tests, are far from perfect. The same is true of the correlations between grades and achievement tests. Personality is positively correlated with grades and achievement test scores. Grades, achievement tests, and IQ capture different aspects of human capabilities.

Figures 1–3 display the predictive power of personality and IQ on grades and scores on achievement tests as measured by the adjusted R^2.[9] The results from the Stella Maris data in Figure 1 indicate that scores on the Raven's Progressive Matrices test explain more of the variance in achievement scores (DAT) than the personality measures. However, personality traits explain a substantial fraction of the variance in the DAT, even when Raven IQ scores are included in regressions. In the Stella Maris data, grades are mostly related to personality traits. Scores on the Raven test do not predict overall grades.

[8]More information about the data sets can be found in Web Appendices 2–5. The study has not been reviewed by an Internal Review Board. There is no need for this because: (1) Three of the four data sets we use are publicly available (BCS, NLSY, MIDUS). (2) The Stella Maris data set has not been submitted to an ethical committee because the research project does not belong to the regimen of the Dutch Act on Medical Research involving Human Subjects (so therefore there is no need for approval of a Medical Ethics committee).

[9]The Web Appendix locations for the source regressions for each figure are given in the notes of each figure.

Figure 2 decomposes achievement tests and grades using data from the British Cohort Study. The results show that IQ and personality measured at age 10 predict scores on various achievement tests at age 10 and age 16, and grades at age 16.

The NLSY data in Figure 3 show that IQ explains more of the variance in AFQT scores and grades than do the only available personality variables—self-esteem and locus of control—but both personality measures are predictive. Note, however, that the measures of personality in the NLSY are only a subset of the wide array of personality traits typically used by psychologists.[10]

The predictive power of personality and IQ for grades and scores on achievement tests is considerably lower in the Stella Maris data compared with the other data sets, which is probably due to the restriction on range in that data set. The sample is constructed from the two highest tracks (out of three possible tracks) at that secondary school.

Some basic patterns emerge across all data sets. Personality predicts grades and scores on achievement tests. IQ is weighted more heavily in predicting achievement scores than in predicting grades. Note that most of the variance in both measures remains unexplained. The reason may be, in part, because of measurement error. But it is also likely that important determinants of these measures are missing in our data sets.

5 Decomposing the contributions of IQ and personality to life outcomes

Using BCS, NLSY, and MIDUS, we determine how much of the variation in numerous important life outcomes is explained by IQ and personality traits. We also consider the relative predictive power of grades and scores on achievement tests compared to IQ. The outcomes studied include wages and measures of health, among other items. We build on the analyses of Borghans et al. (2011), Almlund et al. (2011), and Heckman and Kautz (2012,

[10]See Almlund et al. (2011) for a summary of these measures.

2014).

The results of our analysis of the BCS data plotted in Figure 4 reveal that for wages, years of schooling, the body mass index, number of arrests, and life satisfaction, personality is at least as predictive as IQ.[11] However, the variation explained by IQ and personality is relatively small. Consider, for example, the contribution to explained variance from a regression of log wages on IQ, personality, scores on achievement tests, and grades—reported in various combinations. Column 1 in the first block of columns (corresponding to wages) shows that IQ predicts wages, but the predictive power is small (around 1%). Column 2 shows that self-esteem, locus of control, anti-social behavior, and neuroticism, taken together, are more important determinants of wages. Both IQ and personality remain as important predictors in wage equations when both are included in a regression (column 3). The fourth column shows that achievement has more predictive power than IQ and personality alone. When IQ and personality are also included in a regression (column 5), achievement test scores remain an important predictor of the wage, and IQ and personality also remain as important predictors of the wage. After controlling for scores on achievement tests, IQ loses around 60% of its predictive power. When grades are included, instead of achievement tests, the effect of IQ becomes negligible. A similar pattern arises across the other outcomes studied.

For the NLSY79, Figure 5 parses the contributions of personality and IQ for a set of outcomes. The figure shows that IQ and personality only explain a small portion of the variance for all of the outcomes studied, but that both are important predictors. IQ explains more of the variance than personality for log-wages, any welfare, and physical health at age 40, whereas personality explains more of the variance in mental health at age 40 and whether or not the individual voted in 2006. Achievement tests are better predictors of important life outcomes than IQ.

An analysis of the MIDUS data allows us to consider the predictive power of Big Five personality traits for economic and health outcomes. Figure 6 shows that the Big Five

[11] The adjusted R^2 are displayed in Figures 4–6. The Web Appendix locations of the source regressions are given below each figure.

personality measures in the MIDUS data explain a much larger percent of the variance than IQ for both wage and health outcomes.

The relative importance of IQ and personality measures varies across data sets. This variation is likely driven by differences in the measures used, the choice of measures, the populations considered, and the circumstances under which tests are taken. For example, in the NLSY79, IQ is a better predictor of log wages than personality, but in the BCS and MIDUS data personality measures are better predictors. The better and more comprehensive personality measures in the BCS and MIDUS data compared to those available in the NLSY data likely explain why personality is more predictive of outcomes in those data. The differences may also be driven by the availability of outcomes in each data set as different outcomes most likely place relatively more or less importance on IQ and personality. For example, in both NLSY79 and MIDUS, mental health depends relatively more on personality than physical health.[12]

Despite variation across data sets, consistent patterns emerge. Personality is a powerful predictor for most life outcomes across all data sets. Grades and achievement test scores are more predictive of adult outcomes than IQ. In regression analyses reported in Web Appendix 8, adding grades and test scores to models with IQ and personality produces greater predictive power for the outcomes studied. This larger explained variance is additional evidence that they capture relevant dimensions of human capability not captured by IQ and personality. A general message from our analysis is that further dimensions of achievement remain to be discovered.

6 Conclusions and implications for policy

Cognitive skills predict life outcomes. This paper reinterprets the evidence on the relationship between cognitive skills and a variety of important life outcomes by analyzing the constituent

[12]Errors in the variables can explain some of our evidence. Surprisingly few studies of measurement error in our measures are available. For log wages, measurement error likely explains at most 25% of the variation (Bound et al., 2001).

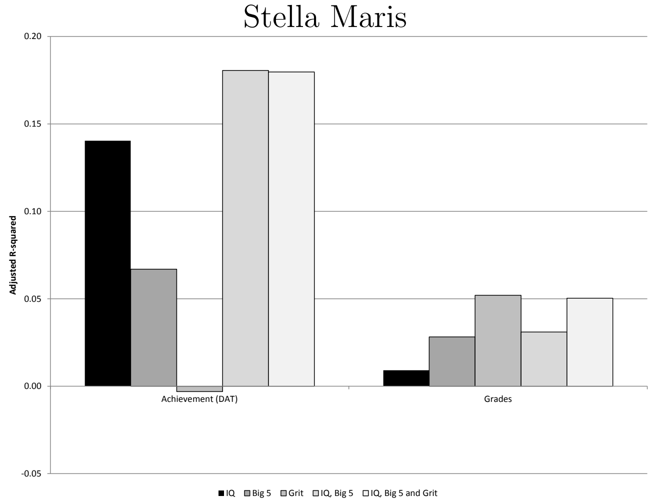
components of widely used proxies for cognitive skills—grades and achievement tests. Measures of personality predict achievement test scores and grades above and beyond IQ scores. Analyses using scores on achievement tests and grades as proxies for IQ conflate the effects of IQ with the effects of personality. Both measures have greater predictive power than IQ and personality alone, because they embody extra dimensions of personality not captured by our measures.

Why do these findings matter? Achievement tests are widely used to measure the traits required for success in school or in life. It is important to know what they measure in order to design effective policy and to use these measures to evaluate schools and teachers.[13] Understanding the sources of differences in the test scores and grades used to explain the black-white achievement gap (Jencks and Phillips, 1998), the male-female wage gap (see, e.g., Bertrand et al., 2010), and other gaps by social class directs attention to what factors might be remediated (see Heckman and Kautz, 2014). For example, personality or non-cognitive skills are more malleable at later ages than IQ, and there are effective adolescent interventions that promote personality but are much less successful in boosting IQ (see Heckman and Mosso, 2014 and Kautz et al., 2014). The predictive power of grades shows the folly of throwing away the information contained in individual teacher assessments in predicting success in life.[14]

[13] See Jackson (2016) on the evidence of teacher effectiveness on personality and its consequences for high school graduation.

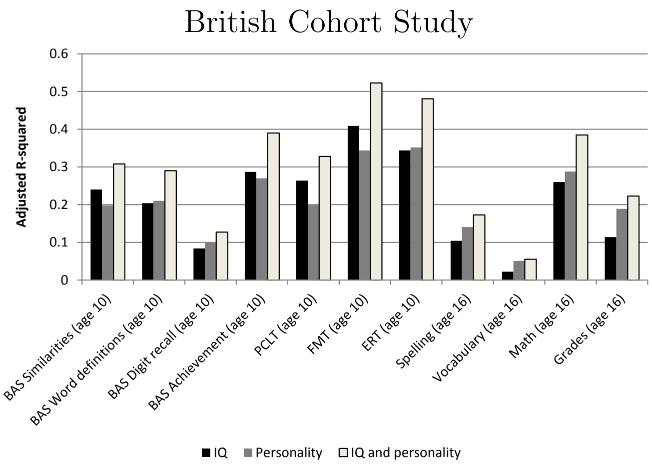
[14] This conclusion echoes the wisdom of Tyler (1940), one of the inventors of the modern achievement test, who recognized the limitations of achievement tests and recognized the value of more comprehensive assessments. His original design for the National Assessment of Educational Progress (NAEP) included more comprehensive measures, including teacher assessments (see Madaus and Stufflebeam, 1989).

Figure 1: Decomposing Achievement Tests and Grades into IQ and Personality



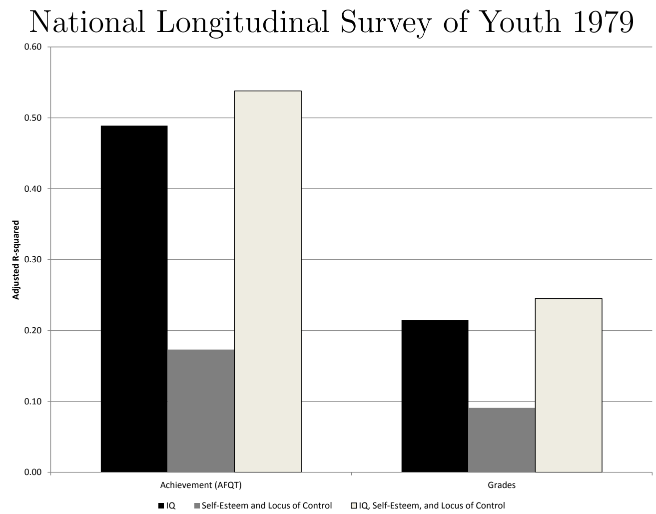
Notes: The Stella Maris data include 347 Dutch high school students aged 15 or 16 in 2008. The figure shows the adjusted R-squareds of two sets of five regressions: (1) DAT/Grades on IQ, (2) DAT/Grades on the Big Five, (3) DAT/Grades on Grit, (4) DAT/Grades on IQ and the Big Five, (5) DAT/Grades on IQ, the Big Five, and Grit. The Big Five (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) from Goldberg (1992) is measured with 10 items per trait. Grit, a measure of perseverance and passion for long-term goals, from Duckworth et al. (2007) is measured with 17 questions. IQ is the principal component of 8 Raven Progressive Matrices. From administrative records, we obtain scores on the Dutch Differential Aptitude Test (DAT) (comparable to the American DAT), an achievement test taken at age 15. Grades are also from administrative records and include the individuals' core subject grade point average at age 13. The curricula of all individuals in the sample are the same at age 13. See Tables 7.1 and 7.2 in the Web Appendix for the regressions supporting these decompositions.

Figure 2: Decomposing Achievement Tests and Grades into IQ and Personality



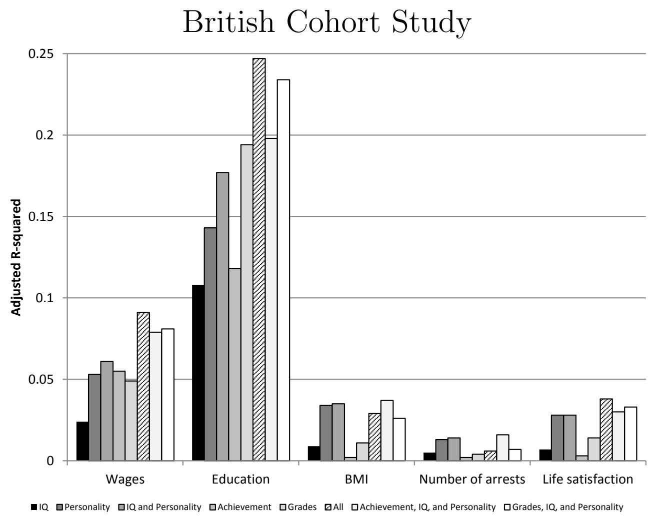
Notes: The British Cohort Study follows a cohort of children born in Britain during one week in April 1970 until 2016. The sample included 17,198 in 1970. The data contain information collected at age 10 on the children's cognitive ability (the Matrices subtest of the British Ability Scales BAS, which is a test similar to the Raven Progressive Matrices test), their personality traits (measures of self-esteem and locus of control based on questions answered by the respondents and measures of disorganized activity, anti-social behavior, neuroticism and introversion based on questions answered by the pupils' teachers) and data from four achievement tests: 1. The BAS achievement test and its three components, 2. The Chess Pictorial Language Comprehension Test, 3. The Friendly Math Test, 4. The Edinburgh Reading Test. At age 16, scores on three other achievement tests are collected: 1. A vocabulary test, 2. A spelling test, and 3. A Math test. Grades is the average grade of 14 subjects at age 16. The Figure shows the adjusted R-squareds of eleven sets of three regressions: (1) Achievement test scores/grades on IQ, (2) Achievement test scores/grades on the personality measures, (3) Achievement test scores/grades on IQ and the personality measures. See Tables 7.3–7.7 in the Web Appendix for the full regressions supporting these decompositions.

Figure 3: Decomposing Achievement Tests and Grades into IQ and Personality



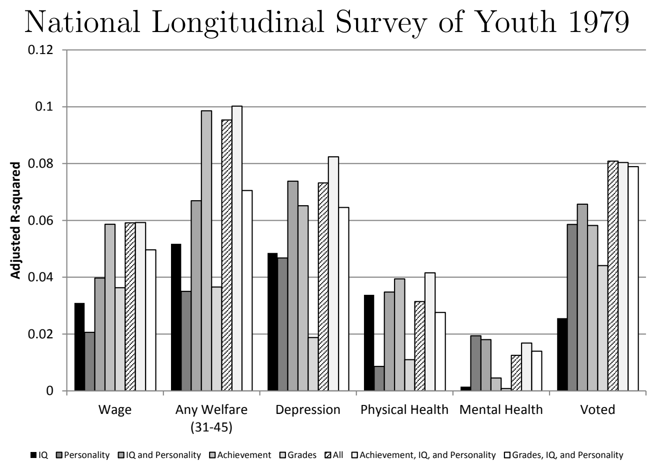
Notes: The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when first surveyed in 1979. The individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. Rotter measures locus of control, was administered in 1979 and is normalized to be mean zero and standard deviation one. Rosenberg measures self-esteem and was administered in 1980. AFQT is measured in 1980. For Rosenberg and Rotter, we use the Item Response Theory (IRT) scores normalized to be mean zero and standard deviation one. AFQT z-scores are constructed from the 1980 percentile score and set to have mean 0 standard deviation 1. IQ and Grades are from high school transcript data. IQ is pooled across several IQ tests using IQ percentiles and then converted into a z-score. Grades are the individual's grade point average from 9th grade and are on a 4 point scale. The sample excludes the military over-sample. Results are shown for the 877 individuals with non-missing IQ, Rotter locus of control, and Rosenberg self-esteem scores. The Figure shows the adjusted R-squareds of two sets of three regressions: (1) Achievement test scores/Grades on IQ, (2) Achievement test scores/Grades on the personality measures, (3) Achievement test scores/Grades on IQ and the personality measures. IQ tests are administered at different ages. Tests taken at early ages may be less predictive. We address this issue in Web Appendix 9. Using IQ tests for more recent surveys (relative to the date of enrollment in the NLSY) does not qualitatively affect our analysis. See Table 7.8 in the Web Appendix for the full regressions supporting these decompositions.

Figure 4: Decomposing Life Outcomes into IQ and Personality



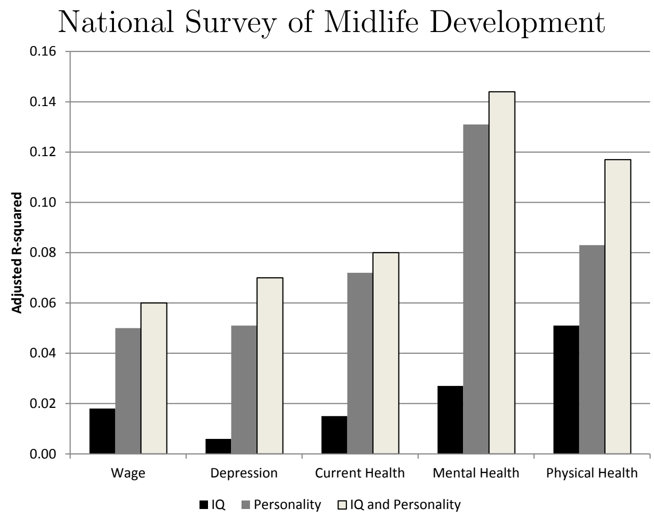
Source: BCS 1970. Notes: See Figure 2. Wages are log Wages at age 38. All other measures are measured at age 34 and standardized to be mean zero and standard deviation one. Education is the nominal age at which a degree is obtained. The Figure shows the adjusted R-squareds of several sets of regressions: (1) Life outcomes on IQ, (2) Life outcomes on the personality measures, (3) Life outcomes on IQ and the personality measures, (4) Life outcomes on achievement (PLCT), (5) Life outcomes on grades, (6) Life outcomes on IQ, Personality, achievement and grades, (7) Life outcomes on achievement, IQ and personality, (8) Life outcomes on grades, IQ and personality. See Tables 8.12–8.16 in the Web Appendix for the full regressions supporting these decompositions.

Figure 5: Decomposing Life Outcomes into IQ and Personality



Notes: Outcomes from the NLSY79. All outcomes are at age 40 unless otherwise noted. Wages are log Wages. Depression is the Center of Epidemiological Studies (CESD) six item depression scale. Physical health is the SF12 self-reported measure of physical health. Mental health is the SF12 self-reported measure of mental health. Voted (2006) is if the individual reports voting in 2006. The Figure shows the adjusted R-squareds of several sets of regressions: (1) Life outcomes on IQ, (2) Life outcomes on the personality measures, (3) Life outcomes on IQ and the personality measures, (4) Life outcomes on achievement, (5) Life outcomes on grades, (6) Life outcomes on IQ, Personality, achievement and grades, (7) Life outcomes on achievement, IQ and personality, (8) Life outcomes on grades, IQ and personality. See Tables 8.1–8.6 in the Web Appendix for the full regressions supporting these decompositions.

Figure 6: Decomposing Life Outcomes into Cognition and Personality



Notes: Data from the National Survey of Midlife Development in the United States 1995-1996 and 2004-2006 (MIDUS). For privacy, income is reported in 42 unique bins in the MIDUS data. We assign individuals the average of their income bin. Sixty-one individuals in the top bin of $200,000 or higher are excluded from the analysis. Cognitive ability is measured by the Brief Test of Adult Cognition by Telephone (BTACT) and personality is measured by the Big Five. Results are restricted to the main sample who were interviewed in both MIDUS I and MIDUS II, have non-missing BTACT and Big Five measures, and are between 30 and 60 years of age during MIDUS II which leaves us with 2,298 observations. All health-related outcomes are from self-reported scales administered during the MIDUS-II follow-up. The Figure shows the adjusted R-squareds of several sets of three regressions: (1) Life outcomes on IQ, (2) Life outcomes on the personality measures, (3) Life outcomes on IQ and the personality measures. See Tables 8.7–8.11 in the Web Appendix for the full regressions supporting these decompositions.

Table 1: Data Analyzed

Datasets	IQ	Achievement Tests	Grades	Personality Measures	Adult Outcomes
Stella Maris (Dutch H.S. students)	✓	✓	✓	✓(Big Five; Grit)	NA
BCS (Children born in one week in 1970 followed until 38)	✓	✓	✓	✓[(1)]	✓
NLSY79 (Prospective survey youth 14–21 in 1979, currently followed)	✓	✓	✓	✓(Self Esteem; Locus of Control)	✓
MIDUS (Survey in adult life, baseline 24–34 in 1995; follow-up 2004–2006)	✓	NA	NA	✓(Big Five)	✓

Note: "NA" denotes "not available." Details on each data set and their measures are provided in Web Appendices 2–5. [(1)] Self esteem, locus of control, disorderly activity, antisocial behavior, introversion, and neuroticism.

Table 2: Correlations (Pearson Correlations)

Correlations	Stella Maris	BCS	NLSY	MIDUS
ρ (IQ, Achievement)	0.378	0.509	0.698	-
ρ (IQ, Grades)	0.112	0.338	0.464	-
ρ (Achievement, Grades)	0.316	0.379	0.610	-
ρ (IQ, Personality)	0.195	0.451	0.291	0.189
ρ (Achievement, Personality)	0.294	0.446	0.410	-
ρ (Grades, Personality)	0.257	0.433	0.305	-

p-values are presented in Web Appendix 6.

References

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

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.

Anastasi, A. and S. Urbina (1997). *Psychological Testing* (7 ed.). Upper Saddle River, NJ: Prentice Hall.

Barton, K., T. E. Dielman, and R. B. Cattell (1972, August). Personality and IQ measures as predictors of school achievement. *Journal of Educational Psychology 63*(4), 398–404.

Bertrand, M., C. Goldin, and L. F. Katz (2010, July). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics 2*(3), 228–255.

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 (2011, August). Identification problems in personality psychology. *Personality and Individual Differences 51*(3: Special Issue on Personality and Economics), 315–320.

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

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.

Duckworth, A. L. and S. M. Carlson (2013). Self-regulation and school success. In B. W. Sokol, F. M. E. Grouzet, and U. Müller (Eds.), *Self-Regulation and Autonomy: Social and Developmental Dimensions of Human Conduct*, Jean Piaget Symposium Series, Chapter 10, pp. 208–230. New York: Cambridge University Press.

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

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

Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2012, May). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology 104*(2), 439–451.

Duckworth, A. L. and M. E. P. Seligman (2005, December). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science 16*(12), 939–944.

Elango, S., A. Hojman, J. L. García, and J. J. Heckman (2016). Early childhood education. Forthcoming, in Moffitt, Robert (ed.), *Means-Tested Transfer Programs in the United States II*. Chicago: University of Chicago Press, 2016.

Goldberg, L. R. (1992, March). The development of markers for the Big-Five factor structure. *Psychological Assessment 4*(1), 26–42.

Green, D. R. (Ed.) (1974). *The Aptitude-Achievement Distinction: Proceedings of the Second*

CTB/McGraw-Hill Conference on Issues in Educational Measurement, Monterey, CA. California Test Bureau: California Test Bureau/McGraw-Hill.

Hanushek, E. A. and L. Woessmann (2008, September). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–668.

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

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

Jackson, C. K. (2016). What do test scores miss? The importance of teacher effects on non-test score outcomes. Under revision, *Journal of Political Economy.*

Jencks, C. and M. Phillips (Eds.) (1998). *The Black-White Test Score Gap*. Washington, DC: Brookings Institution Press.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, Organisation for Economic Co-operation and Development, Paris. Available at https://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf.

Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.), *Educational Measurement*, Chapter 5, pp. 119–158. Washington, DC: American Council on Education.

Madaus, G. F. and D. L. Stufflebeam (Eds.) (1989). *Educational Evaluation: Classic Works of Ralph W. Tyler*. Boston: Kluwer Academic Publishers.

Murnane, R. J., J. B. Willett, and F. Levy (1995, May). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251–266.

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.

Noftle, E. E. and R. W. Robins (2007, July). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology 93*(1), 116–130.

Poropat, A. E. (2009, March). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin 135*(2), 322–338.

Poropat, A. E. (2014, June). A meta-analysis of adult-rated child personality and academic performance in primary education. *British Journal of Educational Psychology 84*(2), 239–252.

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

Tyler, R. W. (1940, September). The place of evaluation in modern education. *The Elementary School Journal 41*(1), 19–27.