What Do Grades and Achievement Tests Measure?*

Lex Borghans
Maastricht University

Bart H.H. Golsteyn
Maastricht University
SOFI, Stockholm University

James J. Heckman
University of Chicago
American Bar Foundation

John Eric Humphries
University of Chicago

December 30, 2015

*This research was supported in part 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 and funded by the Hymen Milgrom Supporting Organization; and the Human Capital and Economic Opportunity Global Working Group, an initiative of the Center for the Economics of Human Development and funded by the Institute for New Economic Thinking. 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 views expressed in this paper are those of the authors and not necessarily those of the funders listed here. The authors received valuable comments among others from 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, the three Spencer conferences at the University of Chicago (2009 and 2010), the Second Conference on Non-Cognitive Skills in Konstanz (2009), the 2009 European Summer Symposium in Labour Economics in Buch am Ammersee, the 2009 meeting of the Association for Research in Personality in Evanston, and the 2009 Society of Labor Economists/European Association of Labour Economists in London. *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 and Research Centre for Education and the Labour Market (ROA), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands and Swedish Institute for Social Research (SOFI), Stockholm University, SE-106 91 Stockholm, Sweden, 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

Grades and achievement tests are widely used as measures of cognition. This paper shows that they are stronger predictors of many important life outcomes than IQ. We examine a variety of sources of evidence on what these measures capture and find that personality contributes substantially to the predictive performance of these measures. This result has important implications for the interpretation of studies using cognition to explain differences in outcomes and for the use of cognitive measures to evaluate the effectiveness of public policies.

Keywords: : IQ, Achievement tests, personality traits

JEL codes: J24, D03

Contents

1 Introduction

Economists and psychologists have established that measures of "intelligence" or "cognition" predict success in a variety of domains of life (see, e.g., 29, 19, 31 and 32). A close look at this evidence shows that different measures of cognitive ability are used by different analysts—grades, IQ, and scores on achievement tests (see the review in 5 and 2). These alternative measures are positively correlated and are often used interchangeably in analyzing "intelligence" (see 31 and 32).

This paper examines these alternative measures and their constituent parts. What concepts do these measures capture? What do they miss? Why is it important to distinguish among these concepts? We establish that, on average, grades and achievement tests are generally better predictors of life outcomes than "pure" measures of IQ. This is because they capture aspects of personality that have been shown to be predictive in their own right. All of the standard measures of "intelligence" or "cognition" are influenced by aspects of personality, albeit to varying degrees, depending on the measure.

Achievement tests were designed to capture "general knowledge" acquired at school and in life (see, e.g., 20, 21). IQ tests were designed to capture "innate aptitudes" rather than acquired knowledge (see 18). The distinction between IQ and achievement tests rests on the crucial notion that aptitudes are fixed while knowledge is not. Yet a large body of scholarly research shows that measures of IQ or aptitude can be altered through interventions (see, e.g., the evidence in 2 and 15). Put differently, all measures of ability require some knowledge in order to gauge the ability which is measured by performance on some task (e.g., taking a test). Knowledge can be acquired. At the same time, greater aptitude facilitates the acquisition of knowledge. Empirically distinguishing knowledge from aptitude is a difficult problem.

All measures of cognition are based on some aspects of acquired knowledge. Aspects of personality affect the acquisition of knowledge (e.g., more motivated persons learn more). At the same time, they directly affect the performance on alternative gauges of cognition for persons with the same basic knowledge (e.g., more conscientious persons take tests more seriously). See Borghans et al. [8].

This paper establishes the following points: (1) Grades, achievement tests and IQ tests all predict life outcomes. The three types of measures are all positively correlated, often strongly so. Higher levels of these measures predict greater success on a number of economic and social dimensions. (2) The predictive power of these measures is ordered for most outcomes: achievement tests and grades are better predictors of most outcomes than are scores on IQ tests. Their greater predictive power is a consequence of the fact that grades and achievement tests capture aspects of personality that have been shown to have independent predictive power of a variety of social and economic outcomes. (3) Personality affects these measures through two channels: (a) by promoting the acquisition of the knowledge that is the basis for the measures and (b) by affecting the performance of persons with the same knowledge on the gauges used to craft the measures.

The paper proceeds as follows. Section 2 gives a brief overview of the literature. Section 3 describes the data. Section 4 decomposes grades and scores on achievement tests into IQ and personality. Section 5 analyzes to what extent other outcomes are predicted by IQ and personality. Section 6 concludes.

2 A brief overview of the literature

In economics, achievement tests such as the Armed Forces Qualification Test (AFQT) are often used as proxies for cognitive ability (See, e.g., 29, 30, and 19). Reviewing the literature, we document 50 papers which use the AFQT as a proxy for intelligence.[1] Many other papers look at other achievement tests or grades as proxies for IQ (see 22, 16, 31 and 32).

The determinants of scores on achievement tests have not been studied in economics. In

[1]The web appendix gives a (non-comprehensive) overview of papers, which use the AFQT as a proxy for intelligence.

psychology, articles have been written about the relationship between (1) IQ and Personality,[2] (2) academic achievement measured by school grades and IQ,[3] and (3) academic achievement measured by school grades and personality.[4,5] However, surprisingly little evidence has been gathered on the extent to which scores on achievement tests are determined by IQ relative to personality.

There are a few exceptions.[6] For example, Duckworth et al. [13] report that self-control (a facet of Big 5 Conscientiousness) and IQ (measured by Raven Matrices) predict scores on the English/Language Arts and Mathematics standardized achievement tests.[7] An early paper by Barton et al. [3] relates the High School Personality Questionnaire and the Culture Fair Intelligence Test to scores on standardized achievement tests. They find that conscientiousness and IQ predict achievement. Duckworth and Carlson [10] provide an overview of the role of self-regulation on school success. In their paper they discuss studies showing how self-regulation is related to standardized achievement tests, course grades, and high school achievement. The paper documents that self-regulation is more predictive of course grades than standardized achievement tests, and suggests this may be why course grades are more predictive of some later-life outcomes than achievement tests.

[2]Duckworth et al. [12] give an overview of this literature. In economics, scores on IQ tests have been related to personality [6]. Segal [37] shows that low conscientious men perform better when they are offered incentives in IQ tests and Borghans et al. [8] 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 spend more time.

[3]Ackerman and Heggestad [1] review the literature.

[4]Poropat [34] and Poropat [35] give an overview of this literature. Poropat [34] concludes that conscientiousness is the largest big five predictor of grades (followed at some distance by openness to experience). Conscientiousness predicts grades almost equally well as intelligence. Poropat [35] evaluates how adolescent measures of the Big 5 predict academic performance – finding that openness and conscientiousness were particularly important. Noftle and Robins [33] investigate the relationship between verbal and mathematic SAT scores and the Big 5. They find that openness to experiences relates to SAT verbal scores.

[5] It is beyond the scope of this paper to review the vast literature in psychology on these relationships. Almlund et al. [2] give a comprehensive overview.

[6]Almlund et al. [2] and Duckworth and Carlson [10] discuss this literature.

[7]Duckworth and Seligman [14] find for a small sample of students that both self-discipline and IQ (the Otis-Lennon School Ability Test Seventh edition) predict performance on the Terranova Second Edition Achievement Test Normal Curve Equivalent. This study reports correlations of the achievement test and IQ and of the achievement test and self-discipline.

3 Data

For our analyses we need data about IQ and personality at a young age, an achievement test at a young age and information about later life outcomes. No data set provides high quality information on all these aspects. We therefore use four data sets. The first data set is from a Dutch high school (Stella Maris). The second is the 1970 British Cohort Study. The third is the U.S. National Longitudinal Study of Youth 1979 (NLSY79). The fourth data set is the National Survey of Midlife Development in the United States (MIDUS).

Data from the Stella Maris High School contain detailed information about achievement, IQ, and personality. They provide information on an achievement test (the Differential Aptitude Test, DAT), a test of cognitive ability (Raven Progressive Matrices), various measures of personality (Big 5, Grit) and measures of school performance (grades).[8,9] Raven Progressive Matrices and the Big 5 are generally considered to be the best measures for fluid intelligence and the diversity in personality among people respectively [5].

The British Cohort Study follows a cohort of children born in Britain during one week in April 1970 until they are 38 years of age.[10] The data contain information collected at age 10 on the children's cognitive ability, some of their personality traits and data from four achievement tests. At age 16, scores on three other achievement tests are collected. In

[8] The sample contains 347 Dutch students, 15 and 16 years of age. (See 6.)

[9] We use the following measures of personality: 50 items to measure the Big 5 (Openness (Cronbach's Alpha=0.73), Conscientiousness (alpha=0.82), Extraversion (alpha=0.86), Agreeableness (alpha=0.79), Neuroticism (alpha=0.81)) from Goldberg [17] and 17 questions to measure Grit, a measure of perseverance and passion for long term goals (alpha=0.700), from Duckworth et al. [11]. We use the principal component of 8 Raven Progressive Matrices as a measure of IQ (alpha=0.62). The Raven matrices are often considered to have the highest loading on g (23). See Jensen [25]. For dissenting views, see Mackintosh and Bennett [27] and Maltby et al. [28]. 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. The DAT and the AFQT are similar in terms of components and the DAT and AFQT correlate highly (0.75). Therefore, conclusions we draw based on the DAT will be instructive about the AFQT as well. **[JJH: We need a reference for this — it's in the GED book]**

[10] The sample included 17,198 babies in April 1970.

addition, these data include many life outcomes.[11] The personality measures are not the standard measures that are used nowadays but cover a richer variety than the NLSY.

The NLSY79 contains Armed Forces Qualifying Test (AFQT) scores, scores on IQ tests,[12] two measures of personality—the Rotter measure of locus of control and Rosenberg measure of self-esteem, grades in high school and many outcomes later in life. The information on personality with only two measures is quite limited, and the IQ information had to be gathered by combining different IQ-tests, that would nowadays not be considered to be the optimal tests for these purposes.

The National Survey of Midlife Development in the United States (MIDUS) is administered in two rounds of interviews. The first was conducted in 1995 and 1996 when respondents were 24-74 years old, and the second was conducted between 2004 and 2006, when respondents were 34-83. MIDUS aimed to be an "interdisciplinary investigation of patterns, predictors, and consequences of midlife development in the areas of physical health, psychological well-being, and social responsibility. The data collection is comprised of four parts.[13]" The data contain psychological measures, background, economic, health, and behavioral data. The survey includes 7,108 respondents (some belonging to a siblings and twins sub-datasets). The survey

[11] Cognitive ability is measured by the Matrices subtest of the British Ability Scales, which is a test similar to the Raven Progressive Matrices test. It contains 28 items. Personality traits include measures of self-esteem (16 items, alpha=0.69) and locus of control (16 items, alpha=0.63) (both based on questions answered by the respondents) and measures of disorganized activity (11 items, alpha=0.93), anti-social behavior (10 items, alpha=0.92), neuroticism (5 items, alpha=0.85) and introversion (5 items, alpha=0.58) (based on questions answered by the pupils' teachers). The achievement tests at age 10 include: 1. the British Ability Scales which include two verbal subscales (Word definitions and Similarities) and two non-verbal subscales (Recall of Digits and Matrices), 2. The Friendly Math Test, 3. The Edinburgh Reading Test and 4. The Chess Pictorial Language Comprehension Test. At age 16, scores on three other achievement tests are collected: 1. A vocabulary test, 2. A spelling test, and 3. The scores on a Math test. Next to this, the data include many life outcomes. We use: 1. Wages (at age 38), 2. Age when left education (asked at age 34), 3. The Body Mass Index (age 34), 4. Number of times been arrested and taken to a policy station (age 34), 5. Satisfaction with life so far (age 34).

[12] The NLSY includes many IQ tests collected from school transcript data for subgroups (the number of respondents is reported in parentheses): California Test of Mental Maturity (599), Lorge-Thorndike Intelligence Test (691), Henmon-Nelson Test of Mental Maturity (201), Kuhlmann-Anderson Intelligence Test (176), Stanford-Binet Intelligence Scale (101), and Wechsler Intelligence Scale for Children (120). The date at which these tests are administered ranges from early childhood to the 12th grade. We use IQ percentile. The advantage of percentile scores is that—in theory—they should be comparable across tests, allowing us to pool test scores from the IQ tests for a much larger sample of test takers.

[13] http://aging.wisc.edu/midus/midus1/index.php

consisted of a telephone survey and mail-based questionnaire. Excluding the oversamples of twins and siblings and excluding respondents who were not interviewed in the second round, the main sample of individuals interviewed in both waves consists of 3,487 individuals. The second round of interviews provides follow-up data of these 3,487 individuals 7 to 11 years later.[14] The analysis in this paper focuses on those 60 and younger during MIDUS II who were part of the main sample. While MIDUS includes measures of personality, cognitive ability, and outcomes, it does not provide an achievement test.

In summary, the NLSY includes the AFQT as an achievement test, which is commonly used in the literature, and is a prospective survey of many life outcomes, but is limited in the quality of both the IQ-measure and measures of personality. The Stella Maris data has a rich set of IQ and personality measures, but only measures short term outcomes of participants in school. BCS has rich IQ measures and long run outcomes of participants, with personality measures that are richer than those in the NLSY but are limited relative to what is available in Stella Maris. Finally, MIDUS provides measures of the Big 5 personality traits, (something absent from almost every other survey with economic outcomes administered in the United States), a measure of cognitive ability, and a number of outcomes. However, it initially interviews participants at later stages of the life cycle.

4 Decomposing grades and the scores on achievement tests into their constituent parts

Figure 1 shows the predictive power of personality and IQ on grades and scores on achievement tests. The results from the Stella Maris data in Panel A indicate that scores on the Raven's

[14] Big 5 personality traits were measured by 30 questions on how well the respondent is described by descriptive adjectives. See Rossi [36] for details. The scale measures the Big 5 personality traits: neuroticism (alpha = 0.74), extraversion (alpha = 0.78), openness (alpha = 0.77), conscientiousness (alpha = 0.58), and agreeableness (alpha = 0.80). Cognitive ability is measured by the Brief Test of Adult Cognition by Telephone (BTACT). The BTACT is the only measure of cognition in MIDUS. The test includes word list recall, delayed word list recall, counting digits backwards, categorical fluency, and number series. While not a formal IQ test, many of the sub-tests are included in some IQ tests. See Tun and Lachman [38] for details.

Progressive Matrices test explain more of the variance in DAT scores than do personality measures. However, personality variables also explain a substantial fraction of the variance in the DAT for both men and women and continue to have predictive power when Raven is included in the same regression. In the Stella Maris data, grades are mostly related to personality traits. The correlations differ for women and men. Grit is an important determinant of GPA for women, while for men, Conscientiousness and Emotional Stability are important. Scores on the Raven test do not predict overall grades for men or women. R-squares increase substantially when personality is added to the regressions. This increase indicates that a much larger part of the variation in GPA is explained by personality than by Raven.

Panel B of Figure 1 decomposes achievement tests using data from the British Cohort Study. The results show that IQ and personality measured at age 10 predict scores on various achievement tests both at age 10 and age 16.

The NLSY data in Panel C show that IQ explains more of the variance in AFQT scores than do Rosenberg and Rotter, but both personality measures are predictive. Decomposing ninth grade grade point average, we find that IQ explains more of the variance in grades than Rosenberg and Rotter, but these personality measures still provide explanatory power above and beyond IQ. However, both personality and IQ measures leave much unaccounted for. It is important to note that our particular measures of personality traits (locus of control and self-esteem) are only a subset of the wide array of personality traits used by psychologists.[15]

Unlike the NLSY79, MIDUS contains measures of Big-5 personality traits, yet it does not contain an achievement test of the quality of AFQT. We rely on MIDUS only for demonstrating the correlations between cognition, personality, and outcomes.

In sum, achievement tests and grades are positively correlated with both IQ and personality. Additionally, personality explains part of achievement test scores beyond that explained by IQ alone.

[15] See Almlund et al. [2] for a summary of these measures.

5 Decomposing the contributions of IQ and personality to life outcomes

Using the BCS, NLSY, and MIDUS, we conduct analyses to determine how much of the variation in several important life outcomes is explained by IQ and personality traits. We build on the analyses of Borghans et al. [7], Almlund et al. [2], Heckman and Kautz [20] and Heckman and Kautz [21]. The array of outcomes includes schooling, health, earnings, and employment.

The BCS data in Figure 2 reveal that for wages, years of schooling, the Body Mass Index, number of arrests and life satisfaction, personality explains at least an equal amount of the variation in life outcomes than does IQ. Notice however that the variation which IQ and personality explain remains low. As a specimen of this array of outcomes, Table 1 shows the wage regressions on IQ, personality traits and various achievement tests. Column 1 shows that IQ predicts the wage. Column 2 shows that self-esteem, locus of control, and neuroticism are determinants of the wage. Both the correlations of IQ and the wage, and of personality and the wage, remain significant when IQ and personality are included in the regression simultaneously (column 3).[16] The fourth column shows that the BAS achievement test has more predictive power than IQ and personality alone. When IQ and personality are also included in the regression (column 5), the BAS achievement test remains an important predictor of the wage, and IQ and personality also remain significant predictors of the wage. In the other columns we run similar regressions using the other three achievement tests as predictors instead of the BAS. We find that achievement remains an important and significant predictor. After controlling for achievement tests, IQ is no longer statistically significant but personality remains a significant predictor of the wage.

Using the NLSY79, Figure 3 parses the contributions of personality and IQ to a larger set of outcomes. The figure shows that IQ and personality only explain a small portion

[16] The R-squareds of column 1, 2 and 3 are displayed in Figure 2.

of the variance in all of the outcomes studied, but that both are important predictors. IQ explains more of the variance for log-wages, any welfare, and physical health at age 40 while personality explains more of the variance in mental health at age 40 and if the individual voted in 2006.

Table 2 shows the full set of regression results for log wages at age 40. The first column includes only IQ, the second column includes only personality, the third column includes IQ and personality, the fourth column includes only AFQT scores. Columns five and six add back in IQ and personality, and column seven includes IQ, personality, and achievement test scores. Table 2 reveals that IQ and Rosenberg are statistically significant predictors of log wages at age 40 when AFQT is not added to the regression. AFQT is a significant predictor of outcomes, even after controlling for personality and IQ. Once AFQT is added, IQ has a small and statistically insignificant coefficient while Rosenberg remains statistically significant. It appears that the achievement test picks up other traits besides IQ and personality which are positively correlated with life outcomes.[17]

Finally, MIDUS allows us to consider the correlation of Big 5 personality traits with economic and health outcomes. Figure 4 demonstrates that the Big 5 personality measures in the MIDUS data explain a much larger percent of the variance than the measure of cognition for both wage and health outcomes. Table 3 shows the coefficients on cognitive ability and the Big 5 personality traits for log wages. We see that cognition is important, but that a larger portion of the variance is explained by personality. Conscientiousness and Openness are positively related to log wages while agreeableness is negatively related. Controlling for personality lowers the coefficient on cognition by almost a third. Results form MIDUS suggest that even more predictive power would be attributed to personality given better measures in our other data sets.

[17] Though both the measures of IQ and personality are limited in the NLSY79, it may be that the achievement test captures aspects of both of these constructs that our limited measures miss.

6 Conclusions and implications for policy

Measured cognitive skills are important determinants of life outcomes. This paper reinterprets the evidence on the relationship between cognitive skills and social outcomes by analyzing the nature 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. Therefore, analyses using scores on achievement tests and grades as proxies for IQ conflate the effects of IQ with the effects of personality. A main conclusion of our paper is that it is important not to interpret predictions of achievement tests and grades with outcomes as solely predictions of IQ with outcomes.

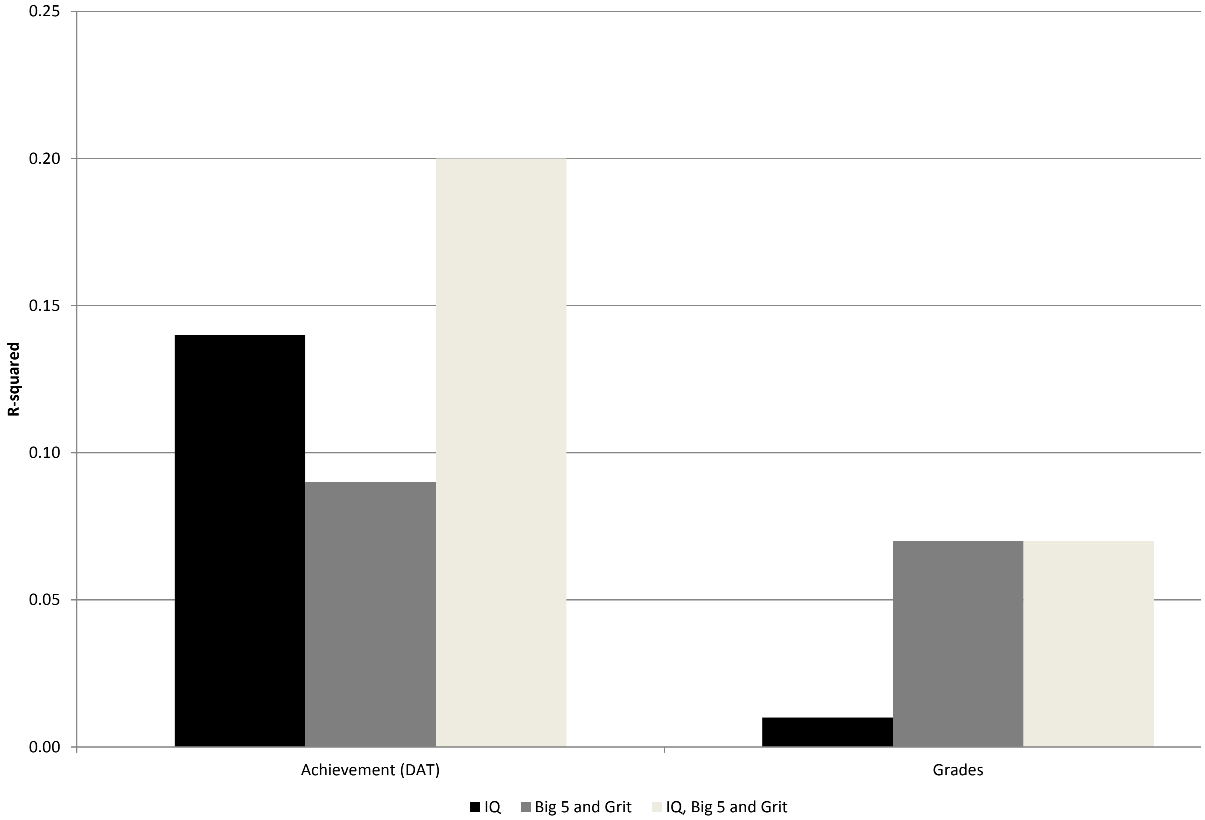
Why does any of this matter? If a test is used to measure the traits required for success in school or in life, then it is important to know what ingredients go into it and how those ingredients are determined to devise wise social policies to promote success. Social policy should recognize and measure the multiple predictors of success. Understanding what constitutes the tests scores and grades used to explain the Black-White achievement gap [24], the male-female wage gap (see, e.g., 4, and 9), and the PISA score and No Child Left Behind test score gaps by social class directs attention to what factors give rise to the gaps and how they might be remediated (see 21).

It is common to measure the performance of children in school by achievement tests. Although achievement tests indirectly measure personality traits as an amalgam, our analysis indicates that personality and cognitive traits should be separated more clearly and controlled for in order to understand test score gaps. For example, the No Child Left Behind Act focuses on improving achievement test scores in math and reading for the lowest part of the distribution. The focus of No Child Left Behind is on improving performance on cognitive tasks. Our analysis shows that personality matters in life, so NCLB's focus likely misdirects attention from personality and other important aspects of education (see, e.g., 26 and 21). Such a focus misleads social policy by focusing on one dimension of performance when many dimensions matter. This affects the design of policies to improve the curriculum.

When Herrnstein and Murray [22] measure the predictive performance of IQ—as measured by AFQT tests—and go on to address the heritable nature of IQ and the difficulty of boosting it using social policy, they confuse many issues. AFQT scores are strongly shaped by personality. Personality traits are malleable (see 21). The dystopic vision of inequality in America by Herrnstein and Murray is reversed by a better understanding of what AQFT actually measures and how its sources can be shaped by social policy.

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

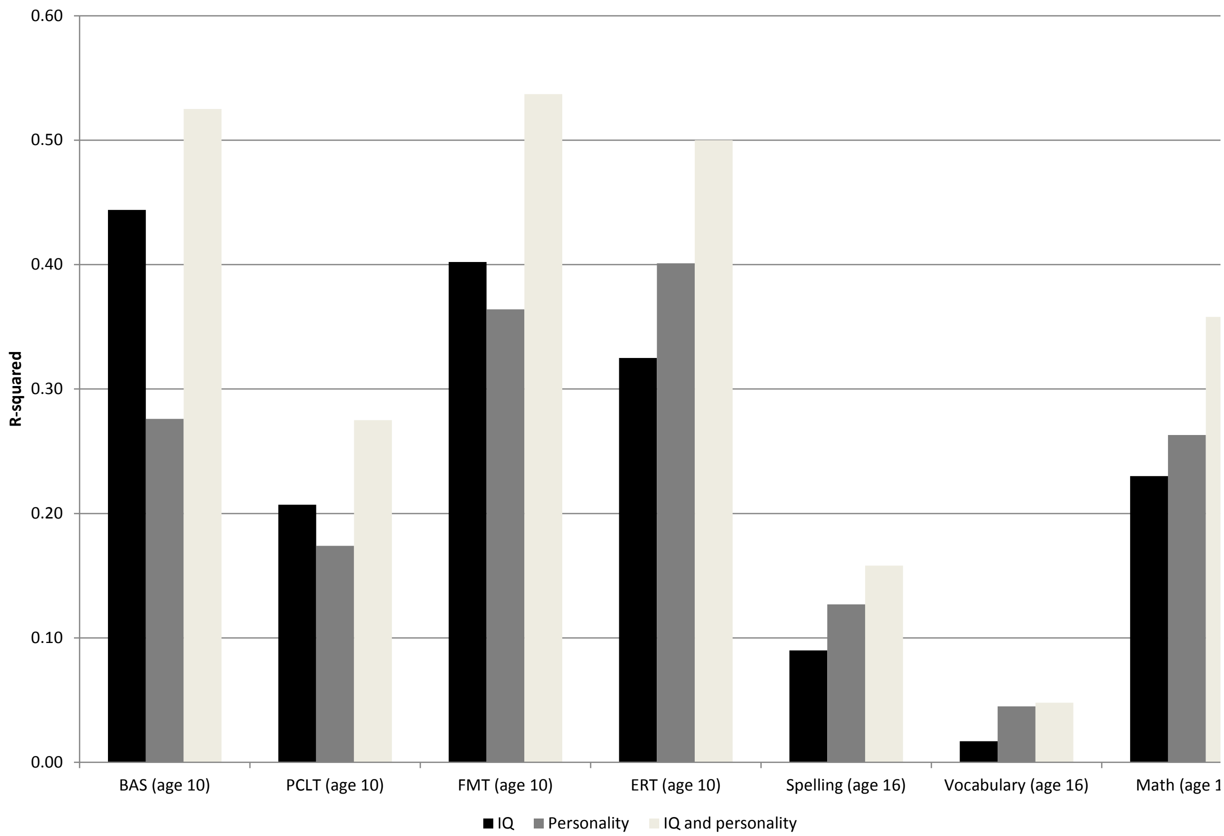
(A) Stella Maris



Notes: The Stella Maris data include 347 Dutch high school students aged 15 or 16 in 2008. The Big 5 (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) from Goldberg [17] is measured with 10 items per trait. Grit, a measure of perseverance and passion for long-term goals, from Duckworth et al. [11] 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.

Figure 1: Decomposing Achievement Tests and Grades into IQ and Personality (cont.)

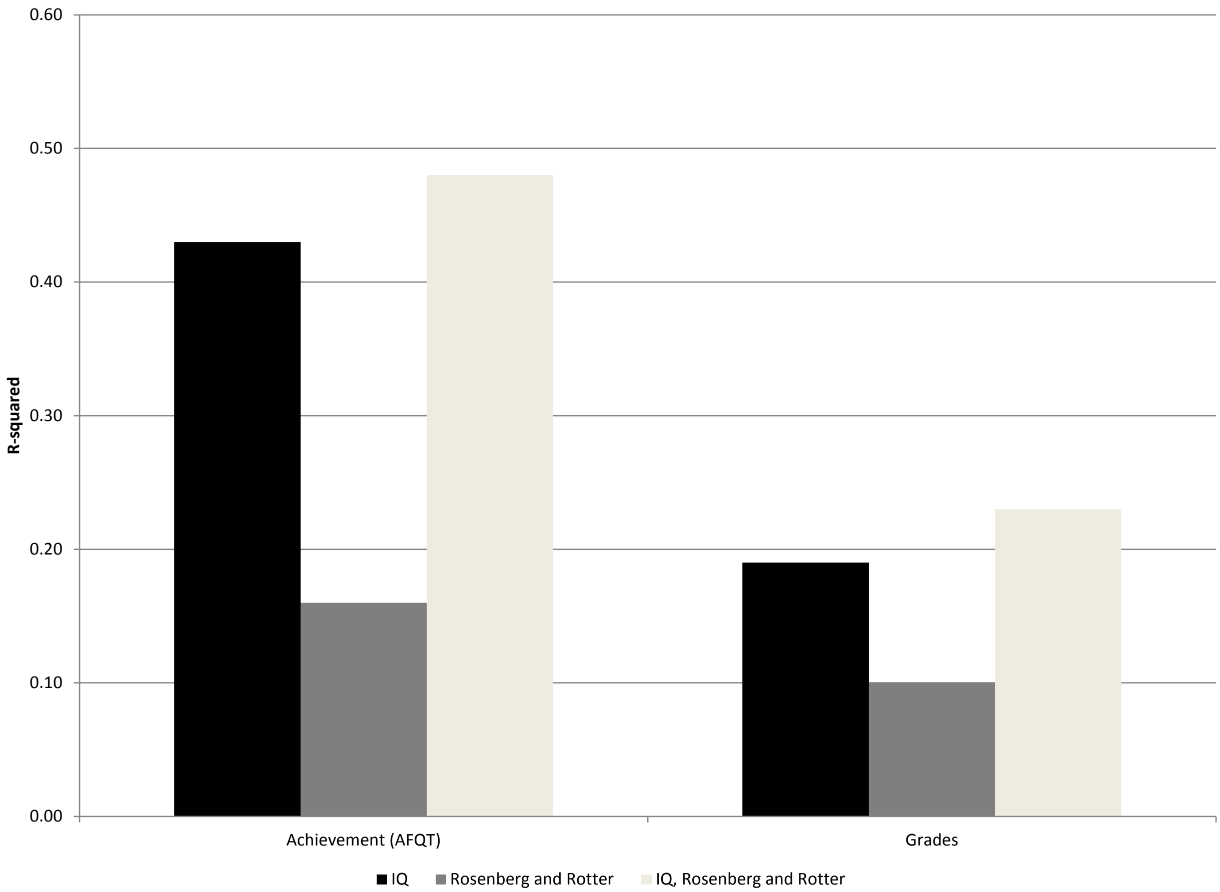
(B) British Cohort Study



Notes: The British Cohort Study follows a cohort of children born in Britain during one week in April 1970 until they are 38 years of age. 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 British Ability Scales which include two verbal subscales (Word definitions and Similarities) and two non-verbal subscales (Recall of Digits and Matrices), 2. The Friendly Math Test, 3. The Edinburgh Reading Test and 4. The Chess Pictorial Language Comprehension Test. At age 16, scores on three other achievement tests are collected: 1. A vocabulary test, 2. A spelling test, and 3. Scores on a Math test.

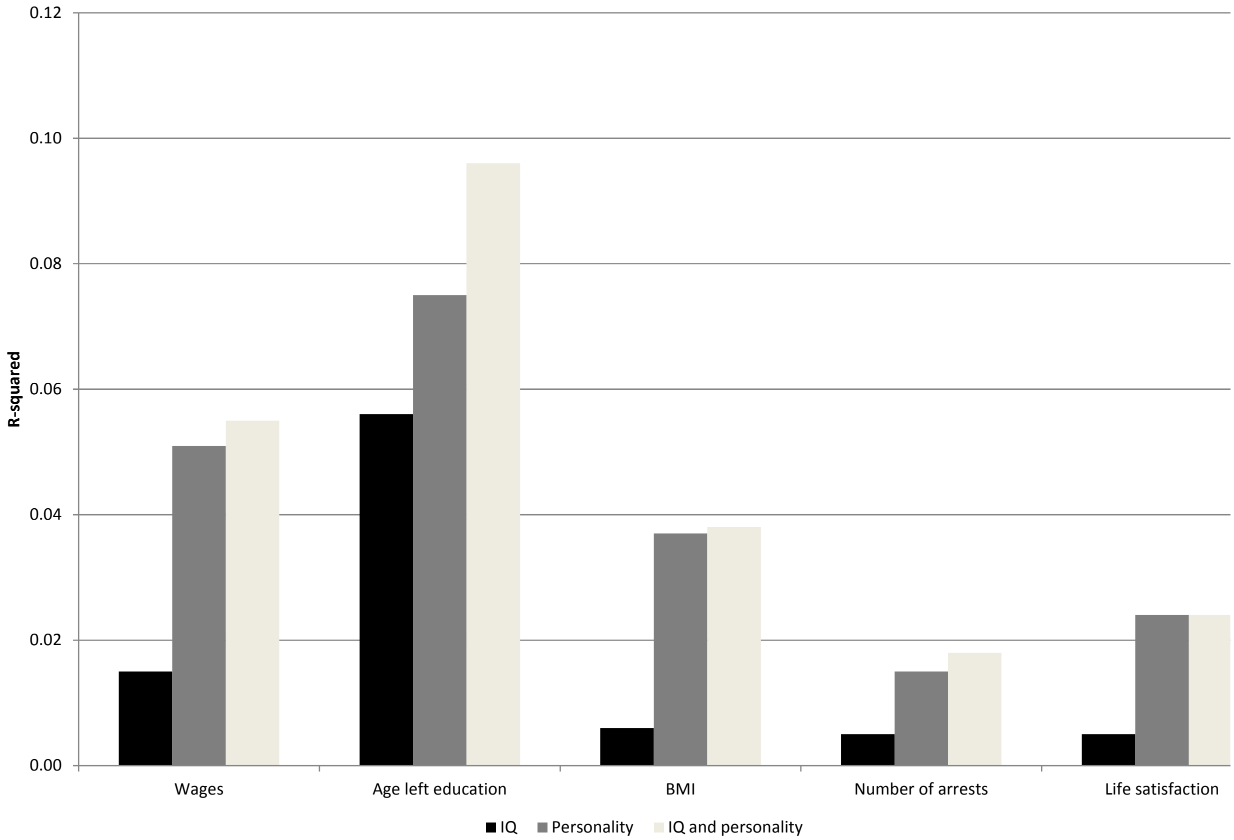
Figure 1: Decomposing Achievement Tests and Grades into IQ and Personality (cont.)

(C) National Longitudinal Survey of Youth 1979



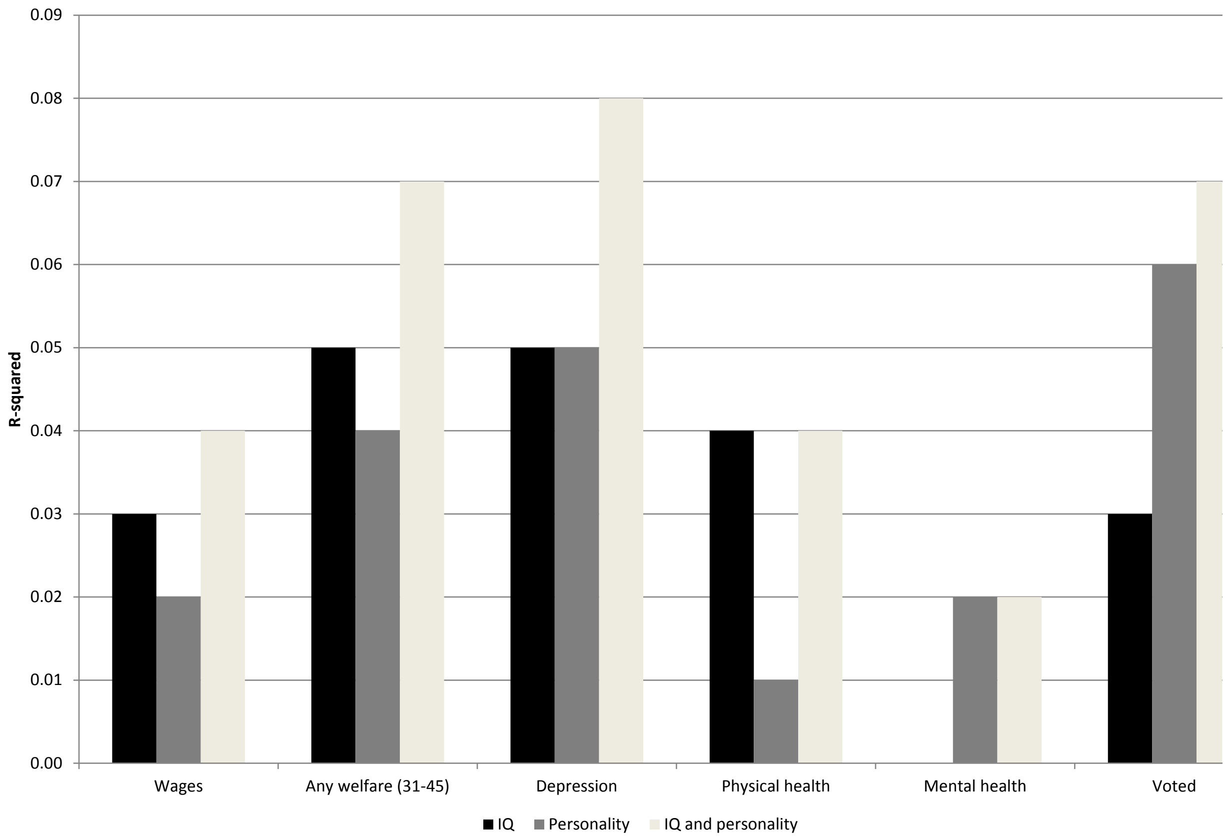
Notes: The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when first surveyed in 1979. The individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. Rotter was administered 1979 and is normalized to be mean zero and standard deviation one. The AFQT and Rosenberg were administered in 1980 and we use the 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. Grades is the individual's grade point average from 9th grade and are on a 4 point scale. Sample excludes the military over-sample.

Figure 2: Decomposing Life Outcomes into IQ and Personality



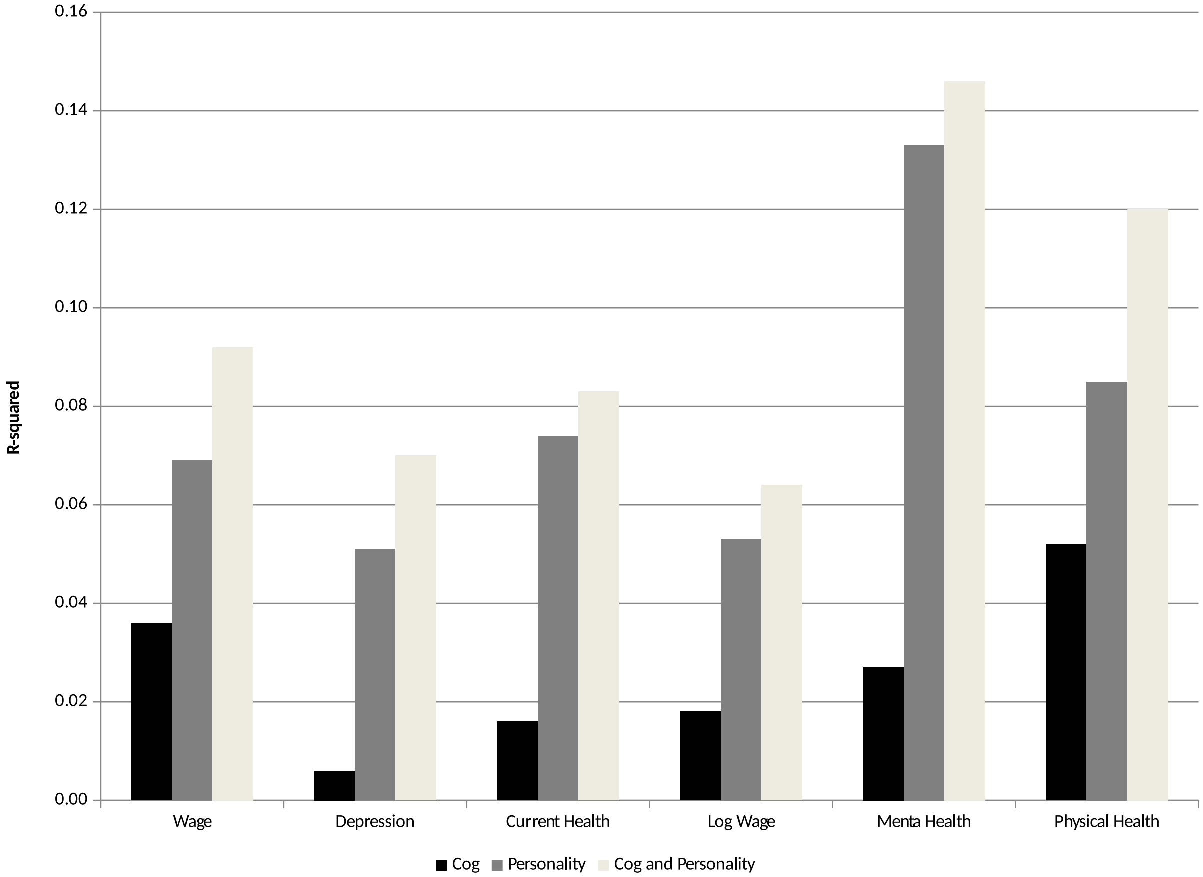
Source: BCS 1970. Notes: See Figure 1B.

Figure 3: Decomposing Life Outcomes into IQ and Personality (NLSY79)



Notes: Outcomes from the NLSY79. All outcomes are at age 40 unless otherwise noted. Depression is the 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.

Figure 4: Decomposing Life Outcomes into Cognition and Personality (MIDUS)



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. All health-related outcomes are from self-reported scales administered during the MIDUS-II follow-up.

Table 1: Log Wage Regressions on Personality, IQ, and Achievement Tests in the BCS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Self-esteem		0.059***	0.058***		0.066***		0.043*		0.046		0.059**
		(0.022)	(0.022)		(0.023)		(0.024)		(0.035)		(0.030)
Locus of control		0.160***	0.146***		0.082***		0.106***		0.071*		0.157***
		(0.024)	(0.024)		(0.026)		(0.027)		(0.039)		(0.036)
Disorganized		-0.001	0.018		0.045		0.026		0.097**		0.049
		(0.028)	(0.029)		(0.030)		(0.031)		(0.048)		(0.042)
Anti-social		0.034	0.031		0.022		0.015		0.019		0.020
		(0.026)	(0.026)		(0.027)		(0.028)		(0.042)		(0.036)
Neurotic		-0.102***	-0.097***		-0.094***		-0.080***		-0.041		-0.071**
		(0.024)	(0.024)		(0.025)		(0.026)		(0.039)		(0.033)
Introvert		-0.013	-0.020		-0.013		-0.038		-0.079**		-0.015
		(0.025)	(0.025)		(0.026)		(0.027)		(0.038)		(0.034)
BAS Matrices (IQ)	0.131***		0.074***		-0.064**		0.023		-0.012		0.036
	(0.021)		(0.022)		(0.027)		(0.026)		(0.040)		(0.033)
BAS Achievement				0.266***	0.254***						
				(0.021)	(0.028)						
PCLT (Achievement)						0.207***	0.149***				
						(0.022)	(0.025)				
FMT (Achievement)								0.387***	0.373***		
								(0.038)	(0.053)		
ERT (Achievement)										0.239***	0.133***
										(0.031)	(0.041)
Constant	0.024	0.003	-0.004	-0.016	-0.018	-0.012	-0.027	0.006	0.010	0.009	0.001
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.022)	(0.022)	(0.033)	(0.034)	(0.030)	(0.029)
Observations	2,589	2,589	2,589	2,362	2,362	2,153	2,153	1,071	1,071	1,339	1,339
R^2	0.015	0.051	0.055	0.066	0.090	0.039	0.064	0.089	0.105	0.043	0.077

Source: BCS 1970.
Notes: See notes below Figure 1B. All independent variables are standardized to mean 0 and standard deviation 1. The sample sizes differ across the columns because the achievement tests were not taken by all respondents. Standard errors in parentheses. *** p <0.01, ** p <0.05, * p <0.1.

Table 2: Log Wages Regressed on IQ, Rosenberg, Rotter, and AFQT in the NLSY79

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IQ	0.173***		0.143***		0.019		0.016
	(0.040)		(0.042)		(0.055)		(0.055)
Rotter		0.165	0.088			−0.007	−0.007
		(0.148)	(0.149)			(0.149)	(0.150)
Rosenberg		0.473***	0.369**			0.275*	0.274*
		(0.147)	(0.149)			(0.150)	(0.150)
AFQT				0.251***	0.237***	0.225***	0.213***
				(0.042)	(0.059)	(0.046)	(0.061)
Constant	10.215***	9.895***	9.985***	10.246***	10.244***	10.108***	10.107***
	(0.041)	(0.100)	(0.102)	(0.041)	(0.041)	(0.107)	(0.107)
Observations	554	554	554	554	554	554	554
R^2	0.033	0.025	0.046	0.060	0.061	0.066	0.066

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 was administered 1979. The AFQT and Rosenberg were administered in 1980. AFQT z-scores are constructed from the 1980 percentile score and set to have a mean of 0 and a standard deviation of 1. IQ and Grades are from high school transcript data. IQ is pooled across several IQ tests using IQ percentiles. Log wage income at age 40 is measured in 2010 dollars. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3: Log Wage Regressed on Cognitive Ability and Personality in MIDUS

	(1)	(2)	(3)
Cognitive Abil.	0.131***		0.102***
	(0.022)		(0.023)
Conscientiousness		0.113***	0.109***
		(0.026)	(0.026)
Extraversion		0.002	0.004
		(0.027)	(0.027)
Neuroticism		-0.074**	-0.065*
		(0.035)	(0.035)
Agreeableness		-0.172***	-0.164***
		(0.023)	(0.023)
Openness		0.114***	0.100***
		(0.027)	(0.027)
Observations	1651	1651	1651
R^2	0.018	0.053	0.064

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. Personality factors are constructed from measures of conscientiousness, extraversion, neuroticism, agreeableness, openness to experience, and agency from the first wave of MIDUS (1995-1996). The cognitive factor is extracted from scores on a series of tests: number series, word list recall, delayed word list recall, categorical fluency, and backwards counting. Factors have been standardized to have a mean of zero and a standard deviation of unity. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

References

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

[2] Almlund, M., A. 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, pp. 1–181. Amsterdam: Elsevier.

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

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

[5] 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.

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

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

[8] 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.

[9] Cattan, S. J. (2014). Psychological traits and the gender wage gap. Unpublished manuscript, Institute for Fiscal Studies.

[10] Duckworth, A. L. and S. M. Carlson (2013). *Self-Regulation and School Success*, Volume 40, Chapter 10, pp. 208. Cambridge University Press.

[11] 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.

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

[13] Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2012). 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.

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

[15] Elango, S., A. Hojman, J. L. García, and J. J. Heckman (2015). Early childhood education. Forthcoming, in Moffitt, Robert (ed.), *Means-tested Transfer Programs in the United States II. Chicago: University of Chicago Press, 2016.*

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

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

[18] Green, D. R. (1974). *The Aptitude-Achievement Distinction.* Monterey: McGraw-Hill.

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

[20] Heckman, J. J. and T. Kautz (2012, August). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464. Adam Smith Lecture.

[21] 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.

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

[23] Huepe, D., M. Roca, N. Salas, A. Canales-Johnson, A. A. Rivera-Rei, L. Zamorano, A. Concepción, F. Manes, and A. Ibañez (2011). Fluid intelligence and psychosocial outcome: From logical problem solving to social adaptation. *PLoS ONE 6*(9), e24858.

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

[25] Jensen, A. R. (1998). *The g factor: the science of mental ability*. Westport, Conn.: Praeger.

[26] Koretz, D. M. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press.

[27] Mackintosh, N. J. and E. S. Bennett (2005). What do Raven's matrices measure? an analysis in terms of sex differences. *Intelligence 33*(6), 663–674.

[28] Maltby, J., L. Day, and A. Macaskill (2010). *Personality, Individual Differences and Intelligence*. Pearson Education.

[29] 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.

[30] Neal, D. A. and W. R. Johnson (1996, October). The role of premarket factors in black-white wage differences. *Journal of Political Economy 104*(5), 869–895.

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

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

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

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

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

[36] Rossi, A. S. (Ed.) (2001). *Caring and Doing for Others: Social Responsibility in the Domains of Family, Work, and Community*. Chicago: University of Chicago Press.

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

[38] Tun, P. A. and M. E. Lachman (2005, January). Brief test of adult cognition by telephone (BTACT). Technical report, Psychology Department, Brandeis University, Waltham, MA.