

# What grades and achievement tests measure

Lex Borghans<sup>a,b</sup>, Bart H. H. Golsteyn<sup>a</sup>, James J. Heckman<sup>c,d,1</sup>, and John Eric Humphries<sup>f</sup>

<sup>a</sup>Department of Economics, Maastricht University, 6200 MD Maastricht, The Netherlands; <sup>b</sup>Research Centre for Education and the Labour Market, Maastricht University, 6200 MD Maastricht, The Netherlands; <sup>c</sup>Department of Economics, The University of Chicago, Chicago, IL 60637; and <sup>d</sup>The American Bar Foundation, Chicago, IL 60611

Contributed by James J. Heckman, September 19, 2016 (sent for review January 22, 2016; reviewed by Armin Falk and Patrick Kyllonen)

**Intelligence quotient (IQ), grades, and scores on achievement tests are widely used as measures of cognition, but the correlations among them are far from perfect. This paper uses a variety of datasets 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 on 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.**

IQ | achievement tests | grades | personality traits

Intelligence quotient (IQ), grades, and scores on achievement tests are widely used as measures of cognition (1, 2) (*SI Appendix, Appendix S1* documents the widespread use of achievement tests as measures of IQ). However, 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 (3–5). They were thought to be more objective and fair 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 (6).

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 (7, 8). Additionally, all measures of ability are based on knowledge as gauged by performance on tasks (e.g., taking a test) (9). 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 (10). In addition, more conscientious people take tests more seriously (11). 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 noncognitive traits that were often deemed to be biased (4, 5).

This paper makes the following points. (i) Grades, scores on achievement tests, and IQ are strongly positively correlated but not perfectly so. This strong correlation gives purchase to the view that the three measures can be used interchangeably. (ii) Grades and scores on achievement tests are differentially influenced by IQ and personality. Grades are more heavily influenced by personality than achievement tests. (iii) All three measures predict a variety of important life outcomes, but scores on achievement tests and grades are better predictors than IQ. (iv) 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 (we make no causal claims in this paper).

## Brief Overview of the Literature

Achievement tests, like the Armed Forces Qualification Test (AFQT), are often used as proxies for cognitive ability (12–14). *SI Appendix, Appendix S1* lists 50 papers that use AFQT scores as proxies for intelligence. Grades are also used as proxies for intelligence (1, 2).

Previous research studies relationships between IQ and personality\*, between grades and IQ (a review of the literature is in ref. 18), and between personality and grades.† Ref. 22 relates the High School Personality Questionnaire and the Culture Fair Intelligence Test to scores on standardized achievement tests and finds that conscientiousness and IQ predict scores on achievement tests. Ref. 23 surveys studies

## Significance

**Grades and scores on achievement tests are widely used as measures of cognition. This paper examines these measures and their constituent parts. We establish that, on average, grades and achievement tests are generally better predictors of life outcomes than “pure” measures of intelligence. The reason is that 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. This result has important implications for the interpretation of studies using scores on achievement tests and grades to explain differences in outcomes and for the use of standard cognitive measures to evaluate the effectiveness of public policies.**

Author contributions: B.H.H.G. and J.J.H. designed research; J.E.H. performed research; B.H.H.G., J.J.H., and J.E.H. analyzed data; and L.B., B.H.H.G., J.J.H., and J.E.H. wrote the paper. The names of the authors are in alphabetical order.

Reviewers: A.F., University of Bonn; and P.K., Educational Testing Service.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. Email: [jjh@uchicago.edu](mailto:jjh@uchicago.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601135113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601135113/-DCSupplemental).

\*Ref. 15 gives an overview of this literature. Scores on IQ tests have been related to personality (16). In related work, ref. 17 shows that less conscientious men perform better when they are offered incentives in IQ tests, and ref. 11 shows that conscientious and emotionally stable people do not spend more time answering IQ questions when rewards are higher, whereas people who score lower on these traits do.

†Refs. 19 and 20 give an overview of this literature. Ref. 19 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. Ref. 20 evaluates how adolescent measures of the Big Five predict academic performance—finding that openness and conscientiousness are particularly important. Ref. 21 investigates the relationship between verbal and mathematical Scholastic Aptitude Test (SAT) scores and the Big Five. It finds that openness to experience relates to SAT verbal scores. Ref. 7 has an extensive review.

**Table 1. Data analyzed**

Datasets	Achievement			Personality measures	Adult outcomes
	IQ	tests	Grades		
Stella Maris (Dutch high school students)	✓	✓	✓	✓(Big Five; grit)	NA
BCS (children born in one week in 1970 followed until 38 y old)	✓	✓	✓	✓*	✓
NLSY79 (prospective survey of youth 14–21 y old in 1979; currently followed)	✓	✓	✓	✓(Self-esteem; locus of control)	✓
MIDUS (survey in adult life; baseline 24–34 y old in 1995; follow-up 2004–2006)	✓	NA	NA	✓(Big Five)	✓

Details on each dataset and their measures are provided in *SI Appendix, Appendices S2–S5*. NA, not available.

\*Self-esteem, locus of control, disorderly activity, antisocial behavior, introversion, and neuroticism.

relating self-regulation and scores on standardized achievement tests, course grades, and high school achievement. It shows that self-regulation is more predictive of course grades than scores on standardized achievement tests and suggests that this may be the reason why course grades are more predictive of certain later-life outcomes than achievement tests. Ref. 24 reports that both self-discipline and IQ predict performance on achievement tests. Ref. 25 reports 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.

## Data

Table 1 summarizes the availability of measures in the four datasets that we analyze.<sup>‡</sup> Although details and point estimates vary and some data contain only partial information, consistent patterns emerge across all four datasets.

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) followed a cohort of children born in one week in April of 1970 until 2016. It has information on grades, IQ, scores on achievement tests, personality, and a variety of adult life outcomes. The National Longitudinal Survey of Youth 1979 (NLSY79) sampled American children aged 14–21 y old in 1979 and followed them ever since that time. It has an achievement test (the AFQT) and scores on different IQ tests across students, which we equate to produce a common IQ score. It has limited measures of 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 y old in 1995–1996 and 34–83 y old in 2004–2006. It has rich data on IQ, personality, and adult outcomes, but lacks information on achievement scores or grades. No single dataset produces definitive evidence. It is the confluence of the evidence across the diverse datasets that justifies the conclusions of this paper.<sup>§</sup>

**Table 2. Correlations (Pearson correlations)**

Correlations	Stella Maris	BCS	NLSY	MIDUS
$\rho$ (IQ, achievement)	0.378	0.509	0.698	—
$\rho$ (IQ, grades)	0.112	0.338	0.464	—
$\rho$ (Achievement, grades)	0.316	0.379	0.610	—
$\rho$ (IQ, personality)	0.195	0.451	0.291	0.189
$\rho$ (Achievement, personality)	0.294	0.446	0.410	—
$\rho$ (Grades, personality)	0.257	0.433	0.305	—

*P* values are presented in *SI Appendix, Appendix S6*.

## 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 datasets. Notice that the correlations between IQ and grades as well as between 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.

Figs. 1, 2, and 3 display the predictive power of personality and IQ on grades and scores on achievement tests as measured by the adjusted  $R^2$ .<sup>¶</sup> The results from the Stella Maris data in Fig. 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.

Fig. 2 decomposes achievement tests and grades using data from the BCS. The results show that IQ and personality measured at age 10 y old predict scores on various achievement tests at ages 10 and 16 y old and grades at age 16 y old.

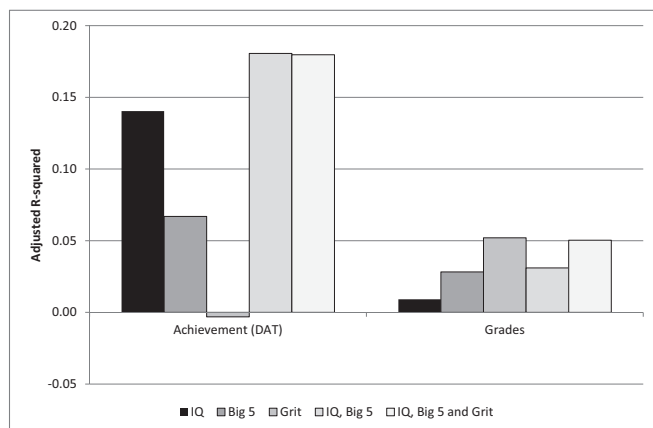
The NLSY data in Fig. 3 show that IQ explains more of the variance in the AFQT scores and grades than the only available personality variables—self-esteem and locus of control—but both personality measures are predictive. Note, however, that the measures of personality in the NLSY are only a subset of the wide array of personality traits typically used by psychologists (ref. 7 has a summary of these measures).

The predictive power of personality and IQ for grades and scores on achievement tests is considerably lower in the Stella

<sup>‡</sup>Across datasets, the survey instruments differ somewhat. The definitions are given in *SI Appendix*.

<sup>§</sup>More information about the datasets can be found in *SI Appendix, Appendices S2–S5*. The study has not been reviewed by an internal review board. There is no need for this because: (i) three of the four datasets we use are publicly available (BCS, NLSY, MIDUS), and (ii) the Stella Maris project does not belong to the regimen of the Dutch Act on medical research involving human subjects. The Stella Maris data were collected at Stella Maris high school with full cooperation of the school. Before the data collection started, all students received a letter with information about the types of questions that were going to be asked. Informed consent was not explicitly asked for because only noninvasive questions were asked. It was mentioned to students that participation was voluntary. In case they did not want to participate, they could indicate this before the data collection started or at any time during the process. One student indicated not to be interested in participating.

<sup>¶</sup>*SI Appendix* locations of the source regressions for Figs. 1, 2, and 3 are given in the notes of each figure.



**Fig. 1.** Decomposing achievement tests and grades into IQ and personality. Stella Maris. The Stella Maris data include 347 Dutch high school students aged 15 or 16 y old in 2008. The figure shows the adjusted  $R^2$  values of two sets of five regressions: DAT/grades on IQ, DAT/grades on the Big Five, DAT/grades on grit, DAT/grades on IQ and the Big Five, and DAT/grades on IQ, the Big Five, and grit. The Big Five (openness, conscientiousness, extraversion, agreeableness, and neuroticism) from ref. 35 are measured with 10 items per trait. Grit, a measure of perseverance and passion for long-term goals, from ref. 36 is measured with 17 questions. IQ is the principal component of eight Raven Progressive Matrices. From administrative records, we obtain scores on the Dutch DAT (comparable with the American DAT), an achievement test taken at age 15 y old. Grades are also from administrative records and include the individuals' core subject grade point average at age 13 y old. The curricula of all individuals in the sample are the same at age 13 y old. *SI Appendix, Tables S7.1 and S7.2* shows the regressions supporting these decompositions.

Maris data compared with the other datasets. 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 datasets, which is probably due to the restriction on range in that dataset. The sample is constructed from the two highest tracks (of three possible tracks) at that secondary school.

Some basic patterns emerge across all datasets. 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. However, it is also likely that important determinants of these measures are missing in our datasets.

### Decomposing the Contributions of IQ and Personality to Life Outcomes

Using the BCS, the NLSY, and the 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 with IQ. The outcomes studied include wages and measures of health among other items. We build on the analyses in refs. 4, 5, 7, and 26.

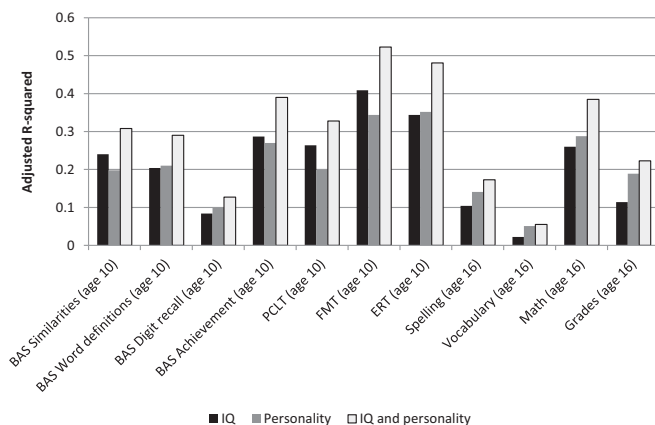
The results of our analysis of the BCS data plotted in Fig. 4 reveal that, for wages, years of schooling, body mass index, number of arrests, and life satisfaction, personality is at least as predictive as IQ.<sup>#</sup> However, the variation explained by IQ and personality is relatively small. Consider, for example, the contribution to explained variance from a regression of log wages on IQ, personality, scores on achievement tests, and grades—reported in

<sup>#</sup>The adjusted  $R^2$  values are displayed in Figs. 4, 5, and 6. *SI Appendix* locations of the source regressions are given below each figure.

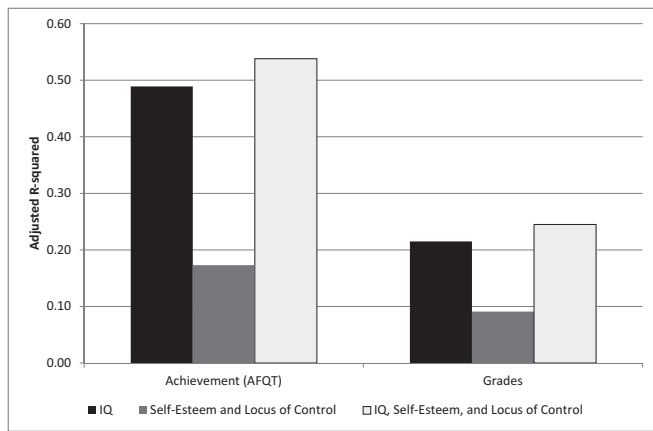
various combinations. Column 1 in Fig. 4 in the first block of columns (corresponding to wages) shows that IQ predicts wages, but the predictive power is small (around 1%). Column 2 in Fig. 4 shows that self-esteem, locus of control, antisocial behavior, and neuroticism, taken together, are more important determinants of wages. Both IQ and personality remain as important predictors in wage equations when both are included in a regression (column 3 in Fig. 4). The fourth column in Fig. 4 shows that achievement has more predictive power than IQ and personality alone. When IQ and personality are also included in a regression (column 5 in Fig. 4), achievement test scores remain an important predictor of wages, and IQ and personality also remain important predictors of wages. 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, Fig. 5 parses the contributions of personality and IQ for a set of outcomes. Fig. 5 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 y old, whereas personality explains more of the variance in mental health at age 40 y old 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 the Big Five personality traits for economic and health outcomes. Fig. 6 shows that the Big Five personality



**Fig. 2.** Decomposing achievement tests and grades into IQ and personality. BCS. The BCS follows a cohort of children born in Britain during one week in April of 1970 until 2016. The sample included 17,198 in 1970. The data contain information collected at age 10 y old 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, antisocial behavior, neuroticism, and introversion based on questions answered by the pupils' teachers), and data from four achievement tests: (i) the BAS achievement test and its three components, (ii) the Chess Pictorial Language Comprehension Test (PCLT), (iii) the Friendly Math Test (FMT), and (iv) the Edinburgh Reading Test (ERT). At age 16 y old, scores on three other achievement tests are collected: (i) a vocabulary test, (ii) a spelling test, and (iii) a math test. Grades are the average grades of 14 subjects at age 16 y old. The figure shows the adjusted  $R^2$  values of 11 sets of three regressions: (i) achievement test scores/grades on IQ, (ii) achievement test scores/grades on the personality measures, and (iii) achievement test scores/grades on IQ and the personality measures. *SI Appendix, Tables S7.3–S7.7* have the full regressions supporting these decompositions.



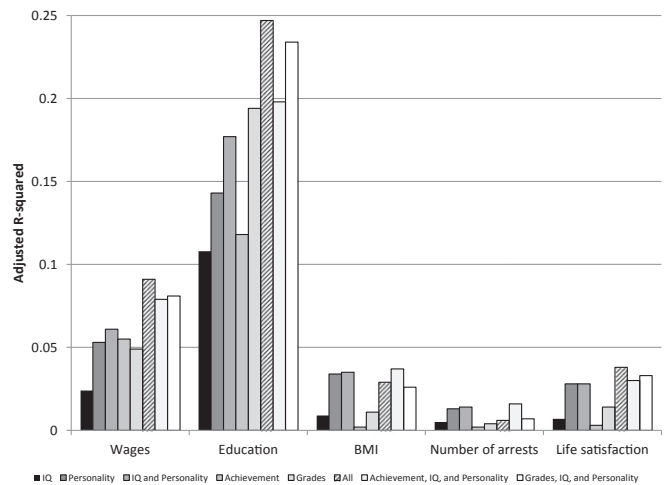
**Fig. 3.** Decomposing achievement tests and grades into IQ and personality. NLSY79. The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14–22 y old when first surveyed in 1979. The individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. Rotten measures locus of control, was administered in 1979, and is normalized to be mean of zero and SD of one. Rosenberg measures self-esteem and was administered in 1980. The AFQT was measured in 1980. For Rosenberg and Rotter, we use the Item Response Theory (IRT) scores normalized to be mean of zero and SD of one. The AFQT z scores are constructed from the 1980 percentile score and set to have mean of zero and SD of one. 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 ninth grade and are on a four-point scale. The sample excludes the military oversample. Results are shown for 877 individuals with nonmissing IQ, Rotter locus of control, and Rosenberg self-esteem scores. The figure shows the adjusted  $R^2$  values of two sets of three regressions: (i) achievement test scores/grades on IQ, (ii) achievement test scores/grades on the personality measures, and (iii) 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 *SI Appendix, Appendix S9*. Using IQ tests for more recent surveys (relative to the date of enrollment in the NLSY) does not qualitatively affect our analysis. *SI Appendix, Table S7.8* shows the full regressions supporting these decompositions.

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

The relative importance of IQ and personality measures varies across datasets. 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 the MIDUS data, personality measures are better predictors. The better and more comprehensive personality measures in the BCS and the MIDUS data compared with 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 dataset, because different outcomes most likely place relatively more or less importance on IQ and personality. For example, in both the NLSY79 and the MIDUS, mental health depends relatively more on personality than physical health.<sup>11</sup>

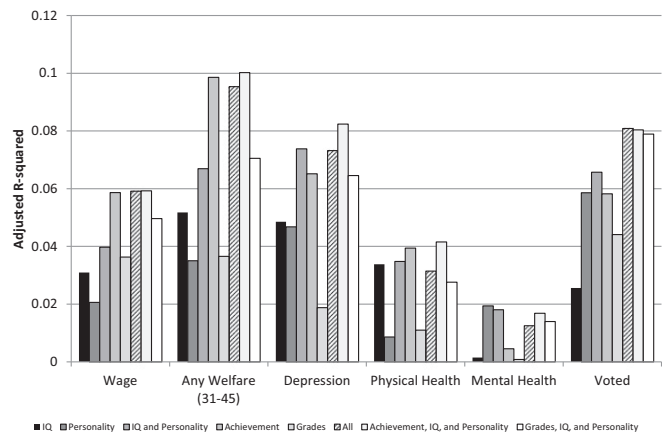
Despite variation across datasets, consistent patterns emerge. Personality is a powerful predictor for most life outcomes across all datasets. Grades and achievement test scores are more predictive of adult outcomes than IQ. In regression analyses reported in *SI Appendix, Appendix S8*, adding grades and test

<sup>11</sup>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 (27).

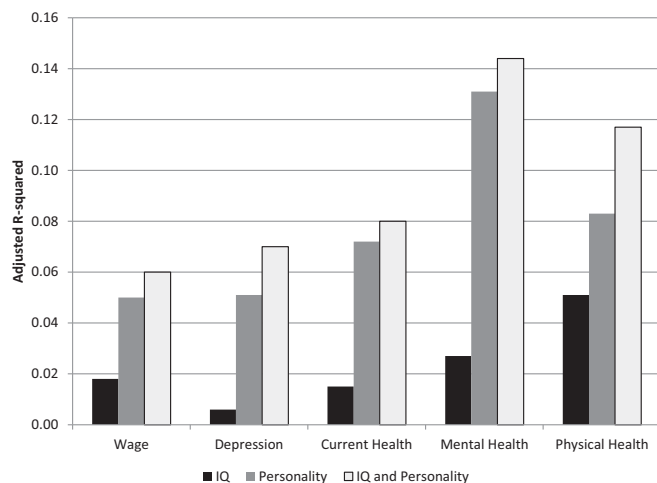


**Fig. 4.** Decomposing life outcomes into IQ and personality. BCS. Source: BCS 1970 (see Fig. 2). Wages are log wages at age 38 y old. All other measures are measured at age 34 y old and standardized to be mean of zero and SD of one. Education is the nominal age at which a degree is obtained. The figure shows the adjusted  $R^2$  values of several sets of regressions: (i) life outcomes on IQ; (ii) life outcomes on the personality measures; (iii) life outcomes on IQ and the personality measures; (iv) life outcomes on achievement (Chess Pictorial Language Comprehension Test); (v) life outcomes on grades; (vi) life outcomes on IQ, personality, achievement, and grades; (vii) life outcomes on achievement, IQ, and personality; and (viii) life outcomes on grades, IQ, and personality. *SI Appendix, Tables S8.12–S8.16* show the full regressions supporting these decompositions. BMI, body mass index.

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



**Fig. 5.** Decomposing life outcomes into IQ and personality. NLSY79. Outcomes from the NLSY79. All outcomes are at age 40 y old 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  $R^2$  values of several sets of regressions: (i) life outcomes on IQ; (ii) life outcomes on the personality measures; (iii) life outcomes on IQ and the personality measures; (iv) life outcomes on achievement; (v) life outcomes on grades; (vi) life outcomes on IQ, personality, achievement, and grades; (vii) life outcomes on achievement, IQ, and personality; and (viii) life outcomes on grades, IQ, and personality. *SI Appendix, Tables S8.1–S8.6* show full regressions supporting these decompositions.



**Fig. 6.** Decomposing life outcomes into cognition and personality. MIDUS. Data from the MIDUS 1995–1996 and 2004–2006. 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 individuals who were interviewed in both MIDUS I and MIDUS II, have nonmissing BTACT and Big Five measures, and were between 30 and 60 y 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  $R^2$  values of several sets of three regressions: (i) life outcomes on IQ, (ii) life outcomes on the personality measures, and (iii) life outcomes on IQ and the personality measures. *SI Appendix, Tables S8.7–S8.11* show the full regressions supporting these decompositions.

general message from our analysis is that additional dimensions of achievement remain to be discovered.

### 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 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 life. It is important to know what they measure to design effective policy and use these measures to evaluate schools and teachers (evidence of teacher effectiveness on personality and its consequences for high school graduation is in ref. 28). Understanding the sources of differences in the test scores and grades used to explain the black–white achievement gap (29), the male–female wage gap (30), and other gaps by social class directs attention to what factors might be remediated (5). For example, personality or noncognitive 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 (31, 32). The predictive power of grades shows the folly of throwing away the information contained in individual teacher assessments when predicting success in life.\*\*

**ACKNOWLEDGMENTS.** We received valuable comments from two referees; Anja Achtziger; Thomas Dohmen; Angela Lee Duckworth; Brent Roberts; Wendy Johnson; Tim Kautz; Anna Sjögren; seminar participants at Institutet för Arbetsmarknads- och Utbildningspolitisk Utvärdering (2010), Swedish Institute for Social Research (2010), Institute for the Study of Labor (2009), Institute for Fiscal Studies (2009), Geary Institute (2009), Tilburg University (2009), and 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. An appendix for this paper can be found at <https://heckman.uchicago.edu/what-do-grades-measure>. This research was supported by the American Bar Foundation; the Pritzker Children's Initiative; the Buffett Early Childhood Fund; NIH Grants National Institute of Child Health and Human Development (NICHD) R37HD065072, NICHD R01HD54702, and National Institute on Aging 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 VIDJ grant from The Netherlands Organization for Scientific Research.

\*\*This conclusion echoes the wisdom of Tyler (33), 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 (34).

- Nisbett RE (2009) *Intelligence and How to Get It: Why Schools and Cultures Count* (Norton, New York).
- Nisbett RE, et al. (2012) Intelligence: New findings and theoretical developments. *Am Psychol* 67(2):130–159.
- Lindquist EF (1951) Preliminary considerations in objective test construction. *Educational Measurement*, ed Lindquist EF (American Council on Education, Washington, DC), pp 119–158.
- Heckman JJ, Kautz T (2012) Hard evidence on soft skills. *Labour Econ* 19(4):451–464.
- Heckman JJ, Kautz T (2014) Fostering and measuring skills: Interventions that improve character and cognition. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, eds Heckman JJ, Humphries JE, Kautz T (Univ of Chicago Press, Chicago), pp 341–430.
- Green DR, ed (1974) *The Aptitude-Achievement Distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement* (California Test Bureau/McGraw-Hill, Monterey, CA).
- Almlund M, Duckworth AL, Heckman JJ, Kautz T (2011) Personality psychology and economics. *Handbook of the Economics of Education*, eds Hanushek EA, Machin S, Wößmann L (Elsevier, Amsterdam), Vol 4, pp 1–181.
- Elango S, Hojman A, Garcia JL, Heckman JJ (2016) Early childhood education. *Means-Tested Transfer Programs in the United States II*, ed Moffitt R (Univ of Chicago Press, Chicago).
- Anastasi A, Urbina S (1997) *Psychological Testing* (Prentice Hall, Upper Saddle River, NJ), 7th Ed.
- Borghans L, Duckworth AL, Heckman JJ, ter Weel B (2008) The economics and psychology of personality traits. *J Hum Resour* 43(4):972–1059.
- Borghans L, Meijers H, ter Weel B (2008) The role of noncognitive skills in explaining cognitive test scores. *Econ Inq* 46(1):2–12.
- Hernstein RJ, Murray CA (1994) *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York).
- Murnane RJ, Willett JB, Levy F (1995) The growing importance of cognitive skills in wage determination. *Rev Econ Stat* 77(2):251–266.
- Hanushek EA, Woessmann L (2008) The role of cognitive skills in economic development. *J Econ Lit* 46(3):607–668.
- Duckworth AL, Quinn PD, Lynam DR, Loeber R, Stouthamer-Loeber M (2011) Role of test motivation in intelligence testing. *Proc Natl Acad Sci USA* 108(19):7716–7720.
- Borghans L, Heckman JJ, Golsteyn BHH, Meijers H (2009) Gender differences in risk aversion and ambiguity aversion. *J Eur Econ Assoc* 7(2–3):649–658.
- Segal C (2012) Working when no one is watching: Motivation, test scores, and economic success. *Manage Sci* 58(8):1438–1457.
- Ackerman PL, Heggestad ED (1997) Intelligence, personality, and interests: Evidence for overlapping traits. *Psychol Bull* 121(2):219–245.

19. Poropat AE (2009) A meta-analysis of the five-factor model of personality and academic performance. *Psychol Bull* 135(2):322–338.
20. Poropat AE (2014) A meta-analysis of adult-rated child personality and academic performance in primary education. *Br J Educ Psychol* 84(Pt 2):239–252.
21. Noftle EE, Robins RW (2007) Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *J Pers Soc Psychol* 93(1):116–130.
22. Barton K, Dielman TE, Cattell RB (1972) Personality and IQ measures as predictors of school achievement. *J Educ Psychol* 63(4):398–404.
23. Duckworth AL, Carlson SM (2013) Self-regulation and school success. *Self-Regulation and Autonomy: Social and Developmental Dimensions of Human Conduct*, Jean Piaget Symposium Series, eds Sokol BW, Grouzet FME, Müller U (Cambridge Univ Press, New York), pp 208–230.
24. Duckworth AL, Seligman MEP (2005) Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol Sci* 16(12):939–944.
25. Duckworth AL, Quinn PD, Tsukayama E (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. *J Educ Psychol* 104(2):439–451.
26. Borghans L, Golsteyn BHH, Heckman JJ, Humphries JE (2011) Identification problems in personality psychology. *Pers Individual Differences* 51(3):315–320.
27. Bound J, Brown C, Mathiowetz N (2001) Measurement error in survey data. *Handbook of Econometrics*, eds Heckman JJ, Leamer EE (Elsevier, Amsterdam), Vol 5, pp 3705–3843.
28. Jackson CK (2016) What do test scores miss? The importance of teacher effects on non-test score outcomes. Working paper (National Bureau of Economic Research, Cambridge, MA). Available at [www.nber.org/papers/w22226](http://www.nber.org/papers/w22226).
29. Jencks C, Phillips M, eds (1998) *The Black-White Test Score Gap* (Brookings Institution Press, Washington, DC).
30. Bertrand M, Goldin C, Katz LF (2010) Dynamics of the gender gap for young professionals in the financial and corporate sectors. *Am Econ J Appl Econ* 2(3):228–255.
31. Heckman JJ, Mosso S (2014) The economics of human development and social mobility. *Annu Rev Econ* 6(1):689–733.
32. Kautz T, Heckman JJ, Diris R, ter Weel B, Borghans L (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).
33. Tyler RW (1940) The place of evaluation in modern education. *Elem Sch J* 41(1):19–27.
34. Madaus GF, Stufflebeam DL, eds (1989) *Educational Evaluation: Classic Works of Ralph W. Tyler* (Kluwer, Boston).
35. Goldberg LR (1992) The development of markers for the Big-Five factor structure. *Psychol Assess* 4(1):26–42.
36. Duckworth AL, Peterson C, Matthews MD, Kelly DR (2007) Grit: Perseverance and passion for long-term goals. *J Pers Soc Psychol* 92(6):1087–1101.

**Web Appendix**

for

**What Grades and Achievement Tests Measure**

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

November 5, 2016

Appendix 1: The literature that interprets the AFQT as an intelligence test.....	5
Appendix 2: Description of the Stella Maris data .....	16
2A. Measures from the experiment.....	17
2B. Data from the administration of the school.....	23
2C. Short Overview of the Schooling System and the Data.....	25
Appendix 3: Description of the British Cohort Study (BCS).....	26
3A. Measure of IQ .....	26
3B. Measures of personality .....	28
3C. Achievement tests .....	29
3D. Grades .....	32
3E. Life outcomes .....	32
Appendix 4: Description of the National Longitudinal Study of Youth, 1979 (NLSY). 33	
4A. Non-cognitive Measurements .....	33
Appendix 5: Description of the National Survey of Midlife Development in the U.S. (MIDUS) .....	37
Appendix 6: Pairwise Pearson Correlations.....	38
Appendix 7: Decomposing Achievement Tests and Grades into IQ and Personality.....	42
Appendix 8: Decomposing the Contributions of IQ and Personality to Life Outcomes.	50
Appendix 9: Using IQ Measures Obtained at Later Ages of Childhood in NLSY79.....	66
10: References .....	67



## Figures

Figure 9.1: NLSY regression of AFQT and Grades on IQ and personality (restricted to those taking IQ tests after 1974 .....	66
--	----

## Tables

Table 1.1: An overview of papers which use the AFQT as a measure for cognitive ability or Intelligence (sorted on date of appearance) .....	5
Table 6.1: Correlations (Pearson Correlations) .....	38
Table 6.2: Correlations Stella Maris data – Full data set .....	39
Table 6.3: Correlations BCS .....	40
Table 6.4: MIDUS Pairwise Pearson Correlations .....	41
Table 6.5: NLYS79 Pairwise Pearson Correlations.....	41
Table 7.1: Decomposing Grades into IQ and Personality – Stella Maris .....	42
Table 7.2: Decomposing DAT Achievement Test into IQ and Personality – Stella Maris .....	43
Table 7.3: Decomposing Achievement Tests into IQ and Personality – BCS.....	44
Table 7.4: Decomposing Achievement Tests into IQ and Personality – BCS (continued) .....	45
Table 7.5: Decomposing Achievement Tests into IQ and Personality – BCS (continued) .....	46
Table 7.6: Decomposing Achievement Tests into IQ and Personality – BCS (continued) .....	47
Table 7.7: Decomposing Grades into IQ and Personality – BCS (continued).....	48
Table 8.1: Log wage income, Age 40 on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY .....	50
Table 8.2: SF12 Mental Health on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY .	51
Table 8.3: SF12 Physical Health on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY	52
Table 8.4: Voted in 2006 election on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY .....	53
Table 8.5: Ever on Welfare between age 31 and 45 on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY.....	54
Table 8.6: Depression on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY .....	55
Table 8.7: Log wage on IQ, Personality and Education – MIDUS .....	56
Table 8.8: Current health on IQ, Personality and Education – MIDUS .....	57
Table 8.9: Physical health on IQ, Personality and Education – MIDUS.....	58
Table 8.10: Mental health on IQ, Personality and Education – MIDUS .....	59
Table 8.11: Depression on IQ, Personality and Education – MIDUS .....	60
Table 8.12: Log wage regressions on Personality, IQ, Achievement, Grades, and Schooling - BCS .....	61
Table 8.13: Regression of Body Mass Index (BMI) on Personality, IQ, Achievement, Grades, and Schooling - BCS .....	62
Table 8.14: Regression of Education on Personality, IQ, Achievement, and Grades – BCS .....	63
Table 8.15: Regressions of Arrests on Personality, IQ, Achievement, Grades, and Schooling – BCS.....	64

Table 8.16: Regressions of Life Satisfaction on Personality, IQ, Achievement, Grades,  
and Schooling – BCS ..... 65

## Appendix 1: The literature that interprets the AFQT as an intelligence test

Table 1.1: An overview of papers which use the AFQT as a measure for cognitive ability or Intelligence (sorted on date of appearance)<sup>1</sup>

Authors	Title	Journal/Book	Quote
Marigee Bacolod, Bernardo S. Blum, William C. Strange	Skills in the city	<i>Journal of Urban Economics</i> 65 (2009) 136–153	The AFQT is designed to measure intelligence.
Jay D. Teachman, Lucky Tedrow	Divorce, Race, and Military Service: More than Equal Pay and Equal Opportunity	<i>Journal of Marriage and Family</i> , Vol. 70, No. 4 (Nov., 2008), pp. 1030-1044	We also control for a number of fixed covariates, including mother's education measured as years of schooling completed as of 1979, intelligence of the respondent measured as his score on the AFQT measured in 1980.
Anne Case, Christina Paxson	Stature and Status: Height, Ability, and Labor Market Outcomes	<i>The Journal of Political Economy</i> , Vol. 116, No. 3 (Jun., 2008), pp. 499-532	Both the NLSY and the Fragile Families samples contain information on the cognitive ability of mothers, in the form of an Armed Forces Qualifying Test (AFQT) score for the mothers in the NLSY.
James J. Heckman, Dimitriy V. Masterov	The Productivity Argument for Investing in Young Children	<i>Review of Agricultural Economics</i> , Vol. 29, No. 3 (Autumn, 2007), pp. 446-493	Figures 8(c) and (d) show that mothers with low cognitive ability provide less cognitive and emotional stimulation for their children.... (c) Average cognitive stimulation score by mother's AFQT decile; and (d) Average emotional stimulation score by mother's AFQT decile.
James J. Heckman, Lance J. Lochner and Petra E. Todd	Earning Functions, Rates of Return and Treatment effects	Chapter 7 in the <i>Handbook of the Economics of Education</i> , 2006	Cognitive ability (as measured by AFQT) is an important determinant of the returns to schooling.

<sup>1</sup>This table is the result of a search in JSTOR on the key words intelligence and AFQT.

Darius Lakdawalla	The Economics of Teacher Quality	<i>Journal of Law and Economics, Vol. 49, No. 1, Symposium: Piracy and File Sharing (Apr., 2006), pp. 285-329</i>	The NLSY 1979 cohort administers to respondents the Armed Forces Qualification Test (AFQT), which is a type of intelligence test.
Pedro Carneiro, James J. Heckman, Dimitriy V. Masterov	Labor Market Discrimination and Racial Differences in Premarket Factors	<i>Journal of Law and Economics, Vol. 48, No. 1 (Apr., 2005), pp. 1-39</i>	Cognitive ability (as measured by the AFQT) of mothers.
Lawrence M. Berger, Jennifer Hill, Jane Waldfogel	Maternity Leave, Early Maternal Employment and Child Health and Development in the US	<i>The Economic Journal, Vol. 115, No. 501, Features (Feb., 2005), pp. F29-F47</i>	Mother's age-adjusted Armed Forces Qualification Test (AFQT) score (a measure of cognitive ability).
Susanne James-Burdumy	The Effect of Maternal Labor Force Participation on Child Development	<i>Journal of Labor Economics, Vol. 23, No. 1 (Jan., 2005), pp. 177-211</i>	Mother's AFQT intelligence score.
David M. Bishai	Does Time Preference Change with Age?	<i>Journal of Population Economics, Vol. 17, No. 4 (Dec., 2004), pp. 583-602</i>	Intelligence as measured by the Armed Forces Qualification Test (AFQT).
Nicola Persico, Andrew Postlewaite, Dan Silverman	The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height	<i>The Journal of Political Economy, Vol. 112, No. 5 (Oct., 2004), pp. 1019-1053</i>	The earliest standardized measure of intellectual ability is the AFQT, an achievement test administered in 1981, when the respondents are 16 or older.
Eric Gould	Inequality and Ability	<i>Labour Economics 12 (2005) 169-189</i>	For the NLSY97 sample, the IQ variable is the age-adjusted AFQT test score which was administered to all members in that sample.
Kai Li, Dale J. Poirier	Bayesian Analysis of an Econometric Model	<i>Journal of Population Economics, Vol. 16, No. 3 (Aug., 2003), pp. 597-625</i>	Variables x7-xi2 capture regional and temporal effects plus the intelligence and family income of the mother (X7=AFQT score/mean of NLSY women).

	of Birth Inputs and Outputs		
Kathy A. Paulson Gjerde	The Existence of Gender-Specific Promotion Standards in the U.S.	<i>Managerial and Decision Economics, Vol. 23, No. 8 (Dec., 2002), pp. 447-459</i>	As discussed at length by Herrnstein and Murray (1994), the AFQT is one of the most highly ‘g-loaded’ mental tests in current use, i.e., measures “general intelligence.” In addition, AFQT scores are highly correlated with a wide range of other mental test scores, lending further support to their use as a measure of IQ. For instance, Scullin et al. (2000), use the AFQT as a proxy for IQ in examining the relationship between general intelligence, educational attainment, and labor market outcomes. Similarly, the AFQT test scores serve as a proxy for general intelligence or ability in the following analysis. Thus, the NLSY data set provides us with the unique opportunity to explore patterns of promotion across multiple occupations and to link these patterns to a measure of general intelligence or ability typically lacking in other data sets. Like all IQ tests, the AFQT
Darlene L. Shearer, Beverly A. Mulvihill, Lorraine V. Klerman, Jan L. Wallander, Mary E. Hovinga, David T. Redden	Association of Early Childbearing and Low Cognitive Ability	<i>Perspectives on Sexual and Reproductive Health, Vol. 34, No. 5 (Sep. - Oct., 2002), pp. 236-243</i>	We based cognitive ability on a composite score derived from the sum of four of the 10 sub-tests: arithmetic reasoning, math knowledge, word knowledge, and paragraph comprehension. Norms for the composite measure—the Armed Forces Qualifications Test (AFQT) score.
Charles Murray	IQ and Income Inequality in a Sample of Sibling Pairs from	<i>The American Economic Review, Vol. 92, No. 2, Papers and Proceedings of</i>	The measure of IQ is the Armed Forces Qualification Test, a highly g-loaded paper-and-pencil test designed for administration to teenage students (g is the general

	Advantaged Family Backgrounds	<i>the One Hundred Fourteenth Annual Meeting of the American Economic Association (May, 2002), pp. 339-343</i>	factor in mental tests). In the NLSY, the average correlation of the AFQT with classic full-scale IQ tests administered to the NLSY sample when they were younger was 0.81, somewhat higher than the usually observed correlations of those IQ tests with each other. The AFQT scores used in the analysis have been normalized separately for each year's birth cohort to a mean of 100 and a standard deviation of 15.
Jane Waldfogel, Wen-Jui Han, Jeanne Brooks-Gunn	The Effects of Early Maternal Employment on Child Cognitive Development	<i>Demography, Vol. 39, No. 2 (May, 2002), pp. 369-392</i>	The AFQT, a measure of cognitive ability that
Samuel Bowles, Herbert Gintis, Melissa Osborne	The Determinants of Earnings: A Behavioral Approach	<i>Journal of Economic Literature, Vol. 39, No. 4 (Dec., 2001), pp. 1137-1176</i>	The Armed Forces Qualification Test (a cognitive test developed to predict vocational success), intelligence tests (e.g. AFQT). Using both the AFQT and a heterogeneous set of IQ (and achievement) test scores, Jencks and Phillips estimate a log earnings function that does not condition on years of schooling. They find an increasing return to a cognitive score between 1983 and 1993. However because they do not include years of schooling as an earnings predictor, it is impossible to say whether their result measures an increased return to schooling or to cognitive performance per se. However, Taber (1997), also using the NLSY (and the AFQT), estimates distinct time trends in the private return to both higher education and to cognitive performance and finds no evidence of an increased earnings effect of the cognitive score for white males between 1979-91. Comparing the covariation of AFQT scores and earnings in black men in the years 1964 and 1980. Jencks and Phillips (1998)

			find that the cognitive score is a stronger predictor of earnings in the latter year. But as in their study just mentioned, these estimates do not condition on the years of schooling, so one cannot tell if these are labor market returns to cognitive skill or a change in the effect of cognitive skill on schooling attainment and/or an increase in the rate of return to schooling for individuals of identical cognitive skill.
Charlie O. Trevor	Interactions among Actual Ease-of-Movement Determinants and Job Satisfaction in the Prediction of Voluntary Turnover	<i>The Academy of Management Journal</i> , Vol. 44, No. 4 (Aug., 2001), pp. 621-638	Cognitive ability was measured in 1980 by giving the NLSY sample the Armed Forces Qualifications Test (AFQT), which is a composite of four quantitative and verbal tests (mathematical knowledge, arithmetic reasoning, paragraph comprehension, and word knowledge). Numerous researchers specifically interested in cognitive ability have used the AFQT composite from the NLSY (e.g., Ganzach, 1998). Additionally, Bock and Moore (1986) reported the AFQT's reliability at over .90, and Gottfredson (1986) has characterized the cognitive ability construct as very stable over time.
Christopher R. Taber	The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?	<i>The Review of Economic Studies</i> , Vol. 68, No. 3 (Jul., 2001), pp. 665-691	I next include the AFQT score in the regression to see how much of this increase can be attributed to changes in the payoff to cognitive ability. ( <i>Although this author also states: Both the AFQT score and some of the scores on other sections of the test will be used as measures of skill that may influence both earnings and schooling decisions.</i> )
James J. Heckman, Yona Rubinstein	The Importance of Noncognitive Skills: Lessons from the GED Testing Program	<i>The American Economic Review</i> , Vol. 91, No. 2, <i>Papers and Proceedings of the Hundred Thirteenth</i>	Recipients are as smart as ordinary high school graduates who do not go on to college, where cognitive ability is measured by an average of cognitive components of the Armed Forces Qualifying Test

		<i>Annual Meeting of the American Economic Association (May, 2001), pp. 145-149</i>	(AFQT) or by the first principle component (g). By these same measures, GED recipients are smarter than other high school dropouts who do not obtain a GED.
Wen-Jui Han, Jane Waldfogel, Jeanne Brooks-Gunn	The Effects of Early Maternal Employment on Later Cognitive and Behavioral Outcomes	<i>Journal of Marriage and Family, Vol. 63, No. 2 (May, 2001), pp. 336-354</i>	If the effects of 1st-year employment are due to the loss of the mothers' cognitive stimulation, then the effects should be more pronounced for children whose mothers have higher cognitive ability themselves. We will test this by comparing the effects of 1st year maternal employment for children whose mothers have differing levels of cognitive ability (as measured by the AFQT).
James Heckman, Edward Vytlačil	Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return of Schooling	<i>The Review of Economics and Statistics, Vol. 83, No. 1 (Feb., 2001), pp. 1-12</i>	In Cawley et al. (1997), we show that there is little difference between general intelligence, AFQT (Armed Forces Qualifying Test), and averages of the ASVAB test of the sort used by Blackburn and Neumark (1993), in terms of explanatory power in log wage regressions.
Charles R. Tittle, Thomas Rotolo	IQ and Stratification: An Empirical Evaluation of Herrnstein and Murray's Social Change Argument	<i>Social Forces, Vol. 79, No. 1 (Sep., 2000), pp. 1-28</i>	Our measure of IQ, the 1980 AFQT score from the NLSY, is the same as that used by Herrnstein and Murray... (Although the authors also state: In using the AFQT for testing the social change argument and alternative interpretations of it, we are not endorsing Herrnstein and Murray's contention that it, or any test, actually measures we are interested in whether those attributes or abilities, whatever they might "intelligence," innate or otherwise. Rather, actually be, that are tapped by measured IQ (AFQT in this instance) predict later status in accordance with supposed trends in "cognitive-enhancing conditions" or "credentialing.")



Judith R. Smith, Jeanne Brooks-Gunn, Pamela K. Klebanov, Kyunghee Lee	Welfare and Work: Complementary Strategies for Low-Income Women?	<i>Journal of Marriage and Family, Vol. 62, No. 3 (Aug., 2000), pp. 808-821</i>	Finally, we do not expect that a mother's limited intellectual abilities, as measured by the Armed Forces Qualifying Test (AFQT), will fully explain the negative effects on child well-being of the mother's lack of employment and her receipt of welfare benefits. (Although these authors also state: Although the AFQT is not an intelligence test, per se, it is highly correlated with IQ and is a widely used measure of adult aptitude.)
David M. Blau	The Effect of Child Care Characteristics on Child Development	<i>The Journal of Human Resources, Vol. 34, No. 4 (Autumn, 1999), pp. 786-822</i>	The AFQT, which is considered to be a measure of intelligence,
Susan K. Lewis, Catherine E. Ross, John Mirowsky	Establishing a Sense of Personal Control in the Transition to Adulthood	<i>Social Forces, Vol. 77, No. 4 (Jun., 1999), pp. 1573-1599</i>	Our measure of cognitive skills is based on test scores. In some cases schools reported the teen's score and percentile rank on one or more intelligence or aptitude tests in 1979. If these were available, we took the mean of the student's percentile ranks on the reported tests. If not, we used the student's percentile rank on the Armed Forces Qualification Test (AFQT), which was administered to all respondents interviewed in 1980
Guang Guo, Leah K. VanWey	Sibship Size and Intellectual Development: Is the Relationship Causal?	<i>American Sociological Review, Vol. 64, No. 2 (Apr., 1999), pp. 169-187</i>	AFQT (a cognitive test)
Yoav Ganzach	Intelligence and Job Satisfaction	<i>The Academy of Management Journal, Vol. 41, No. 5 (Oct., 1998), pp. 526-539</i>	The measure of intelligence was derived from respondents' test scores on the Armed Forces Qualifying Test (AFQT). This test was administered to groups of five to ten respondents between June and October 1980; respondents were compensated, and the overall completion rate was 94 percent. The

			intelligence score was the sum of standardized scores on four tests: arithmetic reasoning, paragraph comprehension, word knowledge, and mathematics knowledge. However, since this score was correlated with age ( $r = .21$ ), I standardized it within each age group to obtain an age-independent measure of intelligence.
Guang Guo	The Timing of the Influences of Cumulative Poverty on Children's Cognitive Ability and Achievement	<i>Social Forces, Vol. 77, No. 1 (Sep., 1998), pp. 257-287</i>	Mother's cognitive ability, AFQT, is a strong ...
Joel Myerson, Mark R. Rank, Fredric Q. Raines, Mark A. Schnitzler	Race and General Cognitive Ability: The Myth of Diminishing Returns to Education	<i>Psychological Science, Vol. 9, No. 2 (Mar., 1998), pp. 139-142</i>	In conducting a test of Herrnstein and Murray's diminishing-returns hypothesis, we used the same data set as they did; the same approach to the selection of respondents, construction of variables, and modeling techniques; and the same test of general cognitive ability, the Armed Forces Qualification Test (AFQT),
Virginia W. Knox	The Effects of Child Support Payments on Developmental Outcomes for Elementary School-Age Children	<i>The Journal of Human Resources, Vol. 31, No. 4 (Autumn, 1996), pp. 816-840</i>	The Armed Forces Qualification Test was administered to NLSY respondents in 1980. It is computed from Arithmetic and Reading sections of the Armed Services Vocational Aptitude Battery, which is an intelligence test used for pre-enlistment screening by the Armed Forces (Baker and Mott, 1989).
Tom Luster, Harriette Pipes McAdoo	Factors Related to the Achievement and Adjustment of Young African American Children	<i>Child Development, Vol. 65, No. 4 (Aug., 1994), pp. 1080-1094</i>	Maternal intellectual ability was measured with the Armed Forces Qualification Test (AFQT), which was administered to the mothers during the 1980 phase of the study.

Donald S. Kenkel, David C. Ribar, Philip J. Cook, Sam Peltzman	Alcohol Consumption and Young Adults' Socioeconomic Status	<i>Brookings Papers on Economic Activity. Microeconomics, Vol. 1994, (1994), pp. 119-175</i>	By results from a standardized intelligence test, the Armed Forces Qualification Test (AFQT)
Toby L. Parcel, Elizabeth G. Menaghan	Early Parental Work, Family Social Capital, and Early Childhood Outcomes	<i>The American Journal of Sociology, Vol. 99, No. 4 (Jan., 1994), pp. 972-1009</i>	Maternal cognitive skills...This concept is measured by the Armed Forces Qualifying Test (AFQT), 1980.
Tom Luster, Robert Boger, Kristi Hannan	Infant Affect and Home Environment	<i>Journal of Marriage and Family, Vol. 55, No. 3 (Aug., 1993), pp. 651-661</i>	Maternal intelligence was assessed in 1980 with the Armed Forces Qualification Test (AFQT),
Mark R. Rosenzweig, Kenneth I. Wolpin	Maternal Expectations and Ex Post Rationalizations: The Usefulness of Survey Information on the Wantedness of Children	<i>The Journal of Human Resources, Vol. 28, No. 2 (Spring, 1993), pp. 205-229</i>	Height and intelligence (AFQT) test scores are available.
Jay Belsky, David Eggebeen	Early and Extensive Maternal Employment and Young Children's Socioemotional Development: Children of the National Longitudinal Survey of Youth	<i>Journal of Marriage and Family, Vol. 53, No. 4 (Nov., 1991), pp. 1083-1098</i>	Mother's intellectual ability was measured by her score on the Armed Forces Qualification Test (AFQT), which was taken by all NLSY participants in 1980.
Marguerite Stevenson Barratt	School-Age Offspring of Adolescent Mothers: Environments and Outcomes	<i>Family Relations, Vol. 40, No. 4 (Oct., 1991), pp. 442- 447</i>	The Armed Forces Qualification Test (AFQT) intelligence measure for the mother was derived from the mother's Profile Scores on the Armed Services Vocational Aptitude Battery (Department of Defense, 1982) administered in 1980.

Kristin A. Moore, Nancy O. Snyder	Cognitive Attainment Among Firstborn Children of Adolescent Mothers	<i>American Sociological Review, Vol. 56, No. 5 (Oct., 1991), pp. 612-624</i>	The mother's cognitive attainments were measured by the Armed Forces Qualifying Test (AFQT).
Douglas K. Detterman, Charlotte G. Andrist	Effect of Instructions on Elementary Cognitive Tasks Sensitive to Individual Differences	<i>The American Journal of Psychology, Vol. 103, No. 3 (Autumn, 1990), pp. 367- 390</i>	General intelligence was measured by using the component score from the first un-rotated factor in a principal component analysis. Input variables included each subject's score on the Armed Forces Qualification Test (AFQT) and 11 subtests of the Armed Services Vocational Aptitude Battery (ASVAB). Intelligence data were available for the 238 Air Force recruits, but not for the 226 National Guard recruits.
Jerry A. Hausman, David A. Wise	Social Experimentation, Truncated Distributions, and Efficient Estimation	<i>Econometrica, Vol. 45, No. 4 (May, 1977), pp. 919-938</i>	Hansen, Weisbrod, and Scanlon [10], however, found intelligence (as measured by the AFQT) to be significantly related to earnings in their sample of "low achievers," who failed to pass the Armed Forces Qualification Test. The effect of education they found to be very small.
Zvi Griliches, William M. Mason	Education, Income, and Ability	<i>The Journal of Political Economy, Vol. 80, No. 3, Part 2: Investment in Education: The Equity- Efficiency Quandary (May - Jun., 1972), pp. S74-S103</i>	Of special interest to us is that a substantial proportion of the veterans' military records contain individual scores on the Armed Forces Qualification Test (AFQT), which we use here in lieu of standard civilian mental ability (IQ) tests.
Courtemanche, Heutel and McAlvanah	Impatience, incentives and obesity	<i>The Economic Journal 125.582 (2015): 1- 31</i>	AFQT percentile proxies for intelligence.
Hani Mansour and Terra McKinnish	Who Marries Differently aged spouses? Ability, education, occupation,	<i>Review of Economics and Statistics 96.3 (2014): 577- 580</i>	Armed Forces Qualifications Test (AFQT)— "cognitive skills measure."

	earnings and appearance		
Daniel P. Mears And Joshua C. Cochran	What is the effect of IQ on offending?	<i>Criminal Justice and Behavior, Vol. 40, No. 11, November 2013, 1280-1300</i>	We use the AFQT measure..., because measures of general intelligence are among the most widely used in the literature on offending.
Pedro de Araujoa, and Stephen Lagosb	Self-esteem, education, and wages revisited	<i>Journal of Economic Psychology</i> 34 (2013): 120-132	Our intelligence variable is derived from Armed Forces Qualifications Test (AFQT) score percentile.
Joseph A. Ritter Lowell J. Taylor	Racial Disparity in Unemployment	<i>The Review of Economics and Statistics</i> 93.1 (2011): 30-42.	The authors use AFQT as cognitive skills measure.
Kevin Lang and Michael Manove	Education and Labor Market Discrimination	<i>American Economic Review</i> 101 (June 2011): 1467–1496	In our empirical analysis, we proxy ability by performance on the Armed Forces Qualification Test (AFQT).
Mohanty	Effects of positive attitude on happiness and wage: Evidence from the US data	<i>Journal of Economic Psychology</i> 30.6 (2009): 884-897.	Intelligence is measured by the Armed Force Qualifications Test (AFQT) score.

## Appendix 2: Description of the Stella Maris data

In June 2008, an experiment was conducted at Stella Maris high school in Meerssen (the Netherlands) (see Borghans et al., 2009). In addition to data collected during the sessions, data from the school administration was also received.

The sample contains 347 Dutch students, 15 and 16 years of age. It provides scores on an achievement test (the Differential Aptitude Test, DAT),<sup>2</sup> a test of cognitive ability (Raven Progressive Matrices), various measures of personality (Big Five, Grit) and measures of school performance (grades). Raven Progressive Matrices and the Big Five are often considered to be the best measures of fluid intelligence and the diversity in personality among people respectively (Borghans et al., 2008).

We use the following measures of personality: 50 items to measure the Big Five (Openness (Cronbach's Alpha=0.73), Conscientiousness (alpha=0.82), Extraversion (alpha=0.86), Agreeableness (alpha=0.79), Neuroticism (alpha=0.81)) from Goldberg (1992) and 17 questions to measure Grit, a measure of perseverance and passion for long term goals (alpha=0.700), from Duckworth et al. (2007).

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 (Huepe et al., 2011). See Jensen (1998). For dissenting views, see Mackintosh and Bennett (2005) and Maltby, Day, and Macaskill (2010).

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 (Kettner, 1976).

Section A describes the measures from the experiment. Section B describes the data from the school administration. Section C gives a short overview of the schooling system in the Netherlands and the timing of the data collection.

---

<sup>2</sup> The DAT is correlated with the ACT, PSAT, and college grades. See Wang (1993) and Omizo (1980).

## 2A. Measures from the experiment

### 2A.1 Measures of psychological traits

#### 2A.1.1 Big Five

50 items for the Big Five from Goldberg (1992):

Conscientiousness ( $\alpha^3=0.822$ )<sup>4</sup>

- I do chores right away
- I leave my things lying around
- I live my life according to schedules
- I neglect my obligations
- I have an eye for details
- I am accurate in my work
- I forget to put things back where they belong
- I am always well prepared
- I often make a mess of things
- I like order

Extraversion ( $\alpha=0.855$ )<sup>5</sup>

- I am the life of the party
- I do not talk a lot
- I feel comfortable around people
- I keep in the background
- I start conversations
- I have little to say
- I talk to a lot of different people at parties
- I do not like to draw attention to myself
- I do not mind being the center of attention
- I am quiet around strangers

Agreeableness ( $\alpha=0.792$ )<sup>6</sup>

- I feel little concern for others
- I am interested in people
- I insult people
- I sympathize with others' feelings
- I am not interested in other people's problems
- I have a soft heart
- I am not really interested in others

---

<sup>3</sup> Cronbach's Alpha is a measure of internal consistency of the trait.

<sup>4</sup> The respondents answer to what extent they agree with the statement. The scale of answers is 1 (not at all) to 5 (totally agree), all items are recoded such that 1=low conscientiousness, 5=high conscientiousness.

<sup>5</sup> The scale of answers is 1-5, all items are now recoded such that 1=low extraversion, 5=high extraversion.

<sup>6</sup> The scale of answers is 1-5, all items are now recoded such that 1=low agreeableness 5=high agreeableness.

I take time out for others  
I feel other people's emotions  
I make people feel at ease

Neuroticism ( $\alpha=0.811$ )<sup>7</sup>

I get stressed out easily  
I am relaxed most of the time  
I worry about things  
I seldom feel blue  
I am easily disturbed  
I get upset easily  
I change my mood a lot  
I have frequent mood swings  
I get irritated easily  
I often feel blue

Openness to experience ( $\alpha=0.725$ )<sup>8</sup>

I have a rich vocabulary  
I have difficulty understanding abstract ideas  
I have a vivid imagination  
I am not interested in abstract ideas  
I have excellent ideas  
I do not have a good imagination  
I am quick to understand things  
I use difficult words  
I spend time reflecting on things  
I am full of ideas

*2A.1.2 Grit*

17 items for Grit and ambition (Ambition, Grit Effort, Grit interest) from the Duckworth et al. (2007) Grit-scale. Some items were rephrased to adjust them for 15 and 16 years old students.<sup>9</sup>  $\alpha=0.700$ .

Ambition ( $\alpha=0.593$ )

I aim to be the best in the world at what I do  
I am ambitious  
Achieving something of lasting importance is the highest goal in life  
I think achievement is overrated  
I am driven to succeed

Grit effort ( $\alpha=0.649$ )

---

<sup>7</sup> The scale of answers is 1-5, all items are now recoded such that 1=low neuroticism 5=high neuroticism.

<sup>8</sup> The scale of answers is 1-5, all items are now recoded such that 1=low openness 5=high openness.

<sup>9</sup> The scale of answers is 1-5, all items are now recoded such that 1=low grit, 5=high grit.



I have overcome setbacks to conquer an important challenge  
Setbacks don't discourage me  
I am a hard worker  
I finish whatever I begin  
I have achieved a goal that took years of work  
I am diligent

Grit interest (0.674)

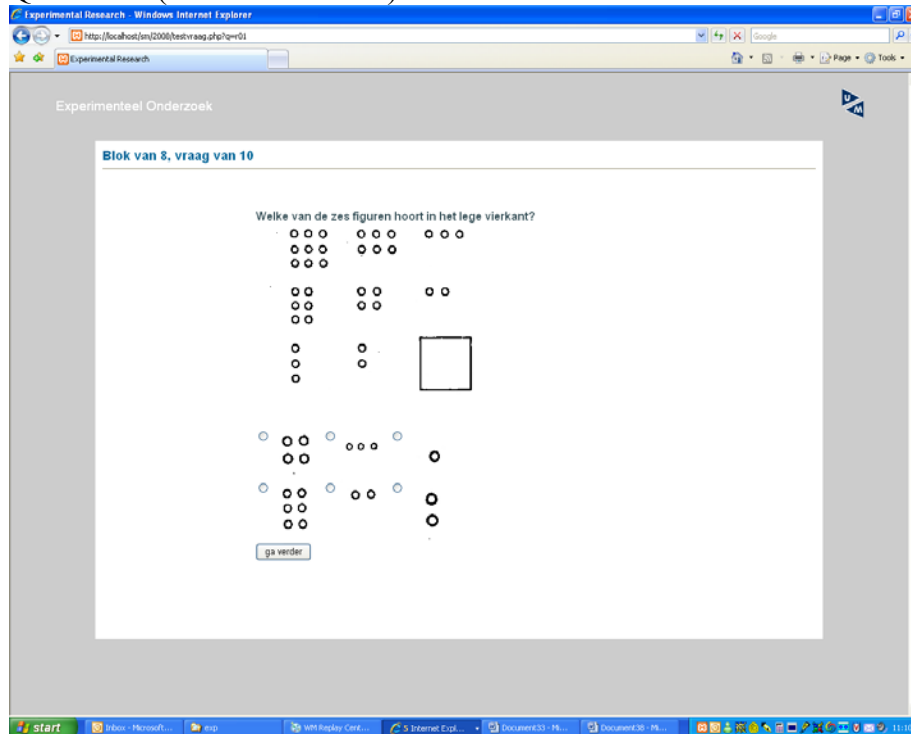
New ideas and projects sometimes distract me from old ones  
My interests change from year to year  
I have been obsessed with a certain project for a short time but later lost interest  
I have difficulty maintaining my focus on projects that take more time than a few months to complete  
I become interested in new pursuits every few months

## 2A.2 Measure of IQ: Raven Progressive Matrices

Eight Raven type questions were asked. We use the sum of the correct answers as a measure for IQ.

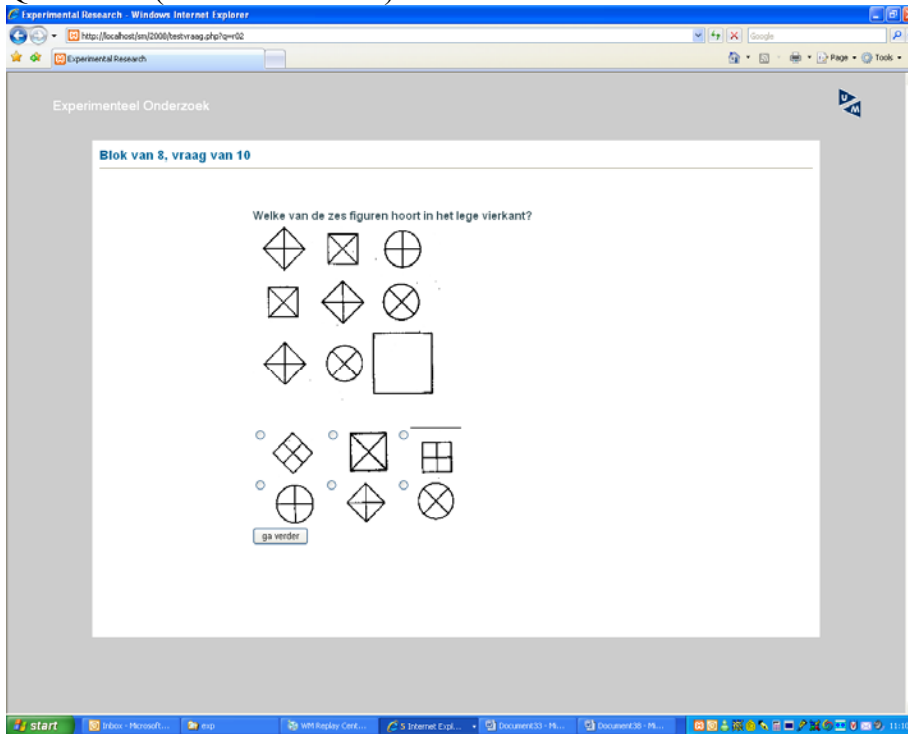
The question reads “Which of the six figures belongs in the empty square?”

Question 1 (correct answer: 3)



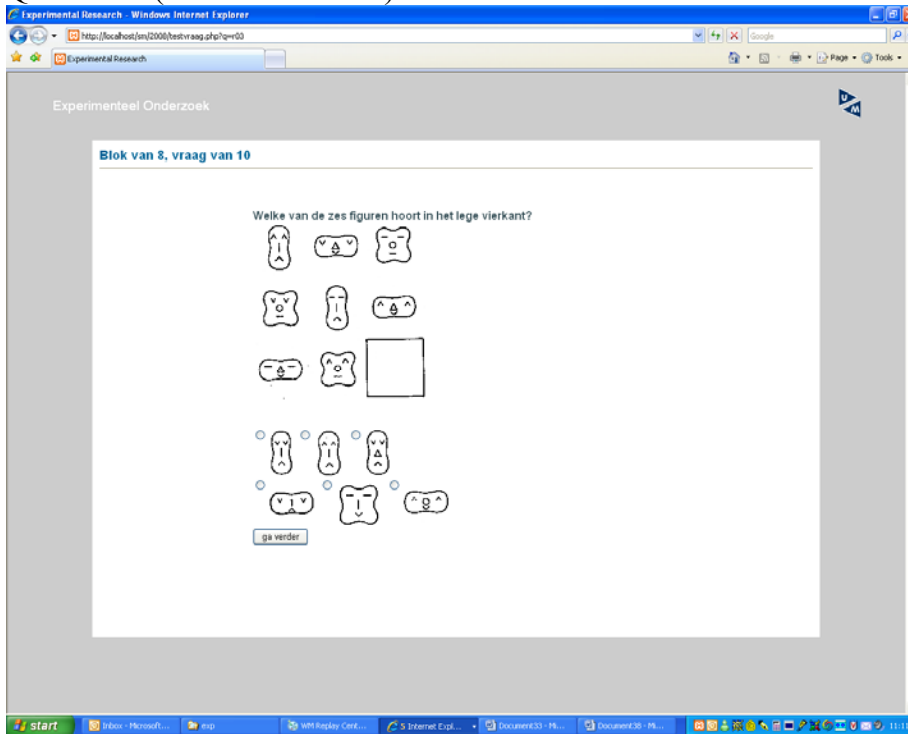
Note: The question reads “Which of the six figures belongs in the empty square?”

Question 2 (correct answer: 3)



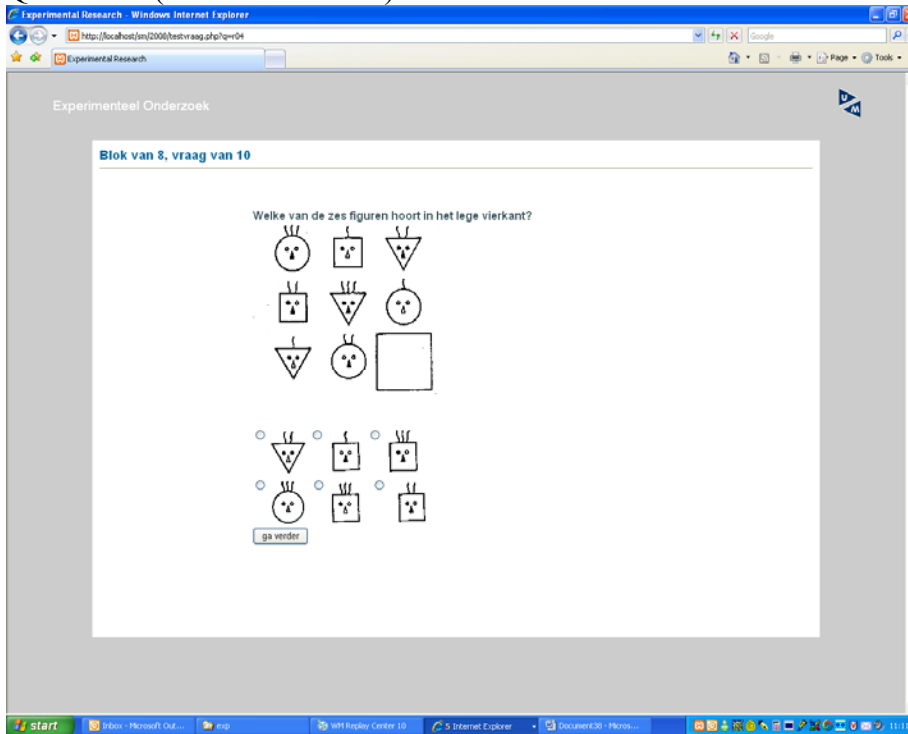
Note: The question reads “Which of the six figures belongs in the empty square?”

Question 3 (correct answer: 1)



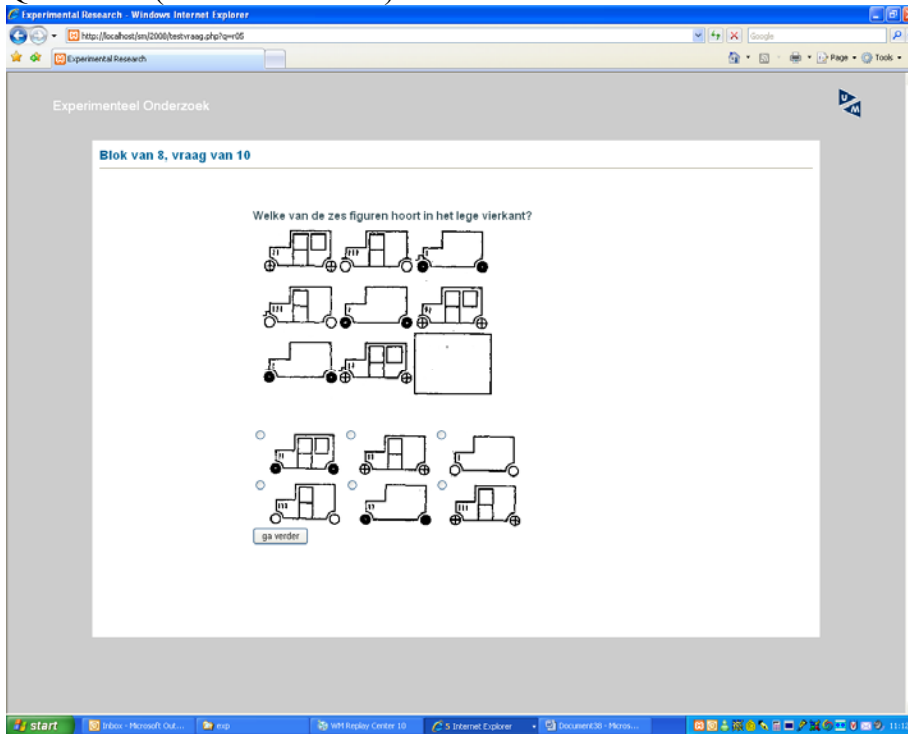
Note: The question reads “Which of the six figures belongs in the empty square?”

### Question 4 (correct answer: 3)



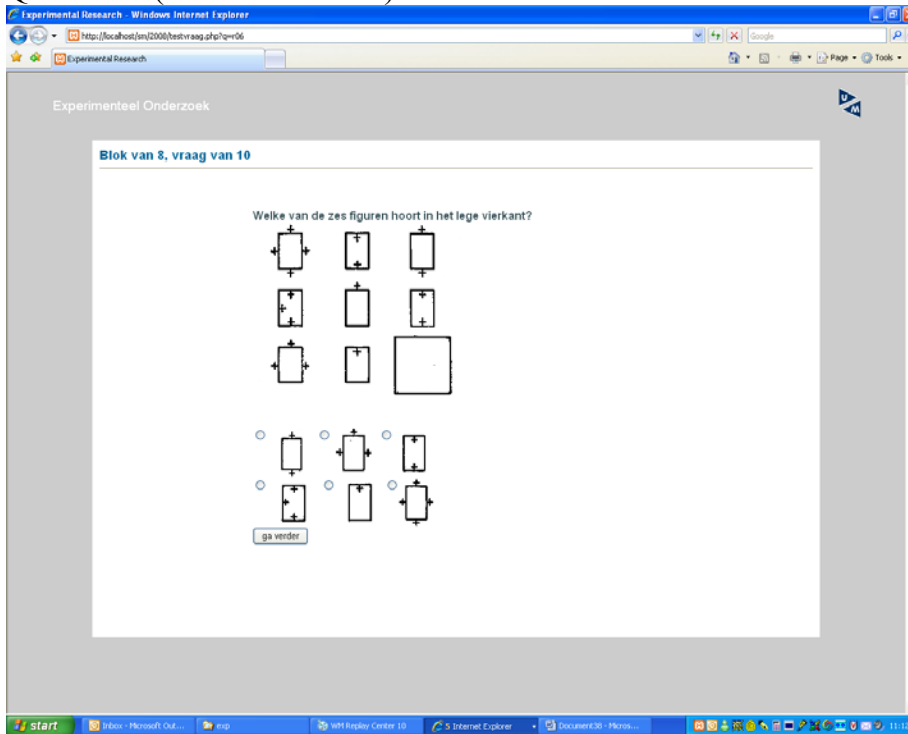
Note: The question reads “Which of the six figures belongs in the empty square?”

### Question 5 (correct answer: 4)



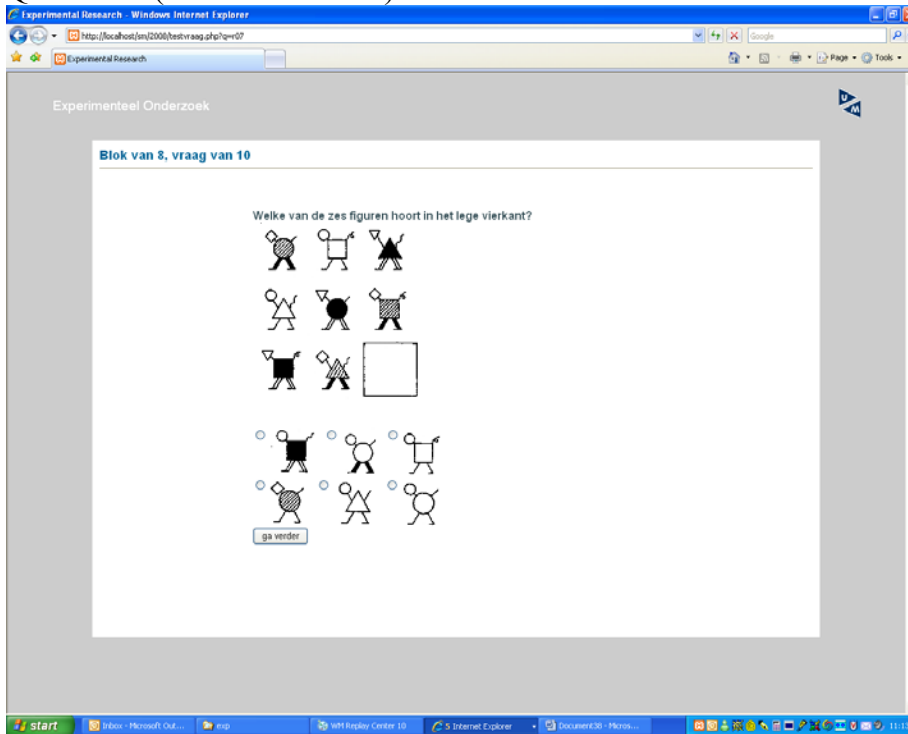
Note: The question reads “Which of the six figures belongs in the empty square?”

### Question 6 (correct answer: 1)



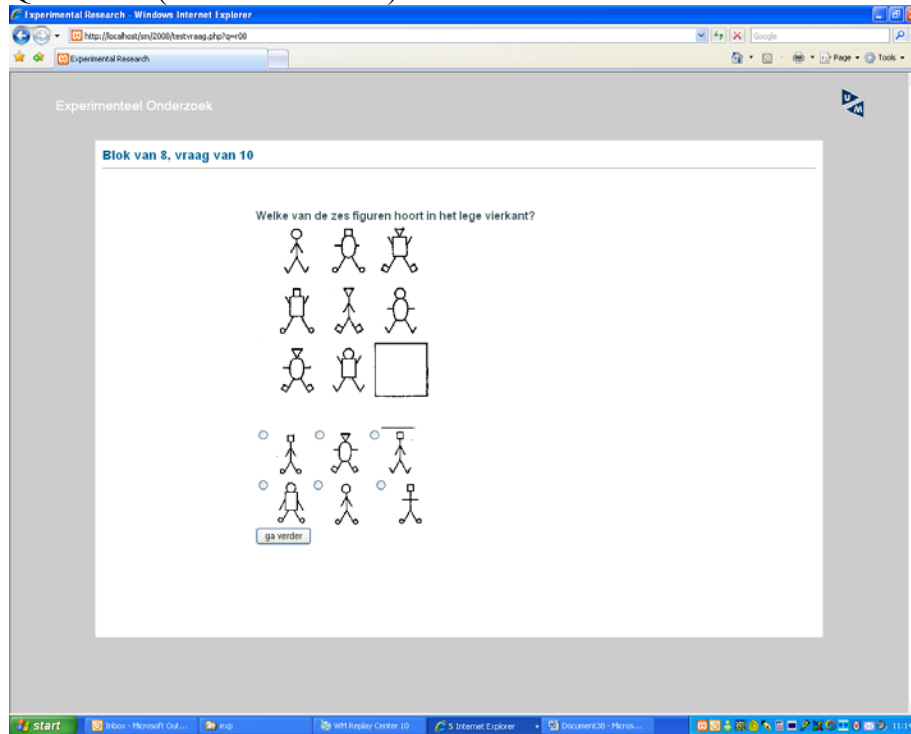
Note: The question reads “Which of the six figures belongs in the empty square?”

### Question 7 (correct answer: 6)



Note: The question reads “Which of the six figures belongs in the empty square?”

## Question 8 (correct answer: 1)



Note: The question reads “Which of the six figures belongs in the empty square?”

## 2B. Data from the administration of the school

### 2B.1 Grades

The data contains grades on all courses (math, physics, chemistry, Dutch, English, German, French, geography, history, etc.) of each year the student attended Stella Maris high school. We use only the grades of courses in the first year the students attended the high school. The reason is that only in this year all students get the same set of courses at the same level of difficulty. After the first year students attend different levels of high school.

### 2B.2 Differential Aptitude Test (DAT)

The DAT is taken as part of a test which informs students about how their abilities and interests fit educations and professions. The DAT measures abilities. Students take the DAT at 14/15 years of age. The data contain each student’s total DAT score for each of the nine subfields. We do not have information on the scores per question within the subfield.

#### SYNONYMS:

The respondent sees a word and has to choose which of 5 options is a synonym (75 tasks, 20 minutes)

**SPELLING:**

The respondent has to choose which of 5 words contains a spelling mistake (40 tasks, 10 minutes)

**LANGUAGE USE:**

The respondent sees a sentence which is cut into 4 pieces. He has to indicate in which of the four pieces a mistake has been made (34 sentences, 10 minutes)

**THINKING WITH WORDS:**

The respondent sees the following: "? word word ?" Then he has to choose from 5 alternatives which pair of words fits on the question marks (40 questions, 20 minutes)

**THINKING WITH NUMBERS:**

Calculus questions and sequence in numbers. The answers are multiple choice with 5 alternatives, the last of which is always "none of the above is correct" (34 questions, 20 minutes)

**SPEED AND ACCURACY:**

The respondent sees a set of 5 combinations of letters and numbers. The respondent has to choose which of 5 answers with the same letters and numbers in another order is not in the first set. (100 questions to practice and 100 questions of actual test, 3 minutes for each set of 100 questions). Example:

One combination in the first list is not in the second list (Correct answer is BA)

BA AC AD AE AF

AC AE AF AB AD

**THINKING WITH FIGURES:**

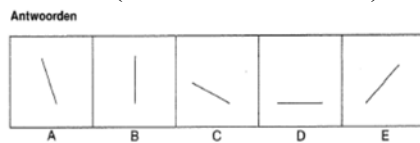
The respondent sees a series of 4 figures. For each series, he has to choose which one of 5 alternatives is the logic continuation of the series. (30 series, 15 minutes).

Example:

Which figure fits best in the following series:



Answers (correct answer is D):



**THREE-DIMENSIONAL UNDERSTANDING:**

The respondent has to choose which one of 4 alternative objects fits best to a given folded pattern. (40 questions, 15 minutes)

**PRACTICAL INSIGHT:**

The respondent has to choose how a practical problem can be solved. In each task three answer categories are given. The questions are about mechanical tools or simple physical principles. (50 questions, 10 minutes)

## 2C. Short Overview of the Schooling System and the Data

### LAST YEAR PRIMARY SCHOOL

- In the last year in primary school (age 12), students do the CITO test. Based on the results of this test and previous experience of the primary school with the student, the primary school advises students to go to the lowest, middle or upper level high school.

### FIRST YEAR HIGH SCHOOL

- After primary school students go to high school (age 13).
- The Stella Maris high school has two levels: the middle (HAVO) and upper (VWO) level. There are no students from the lowest (VMBO) high school level in this school.
- In the first year, all students in the school (so the students from the middle and upper track) attend the same courses at the same level.
- At the end of the first year, the high school decides whether the students should attend the middle or upper level. A very small group has to leave the school to go to the lowest level.
- The data contain the grades of the students for all courses and the level to which they were assigned (middle or upper).

### THIRD YEAR HIGH SCHOOL

- All students did the Differential Aptitude Test (DAT) in this year.
- Half of the sample participated in the experiment while in year 3. In the experiment Raven, Big 5, and Grit are measured.
- In the third year, students have to choose which profile they are going to follow in the fourth year. There are 4 profiles: 1. culture and society, 2. economics and society, 3. nature and health, 4. nature and technology.

### FOURTH YEAR HIGH SCHOOL

- The other half of the sample participated in the experiment while in year 4.

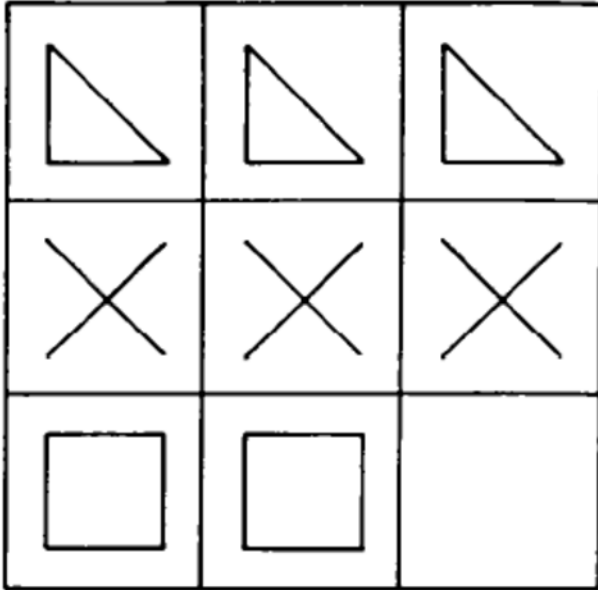




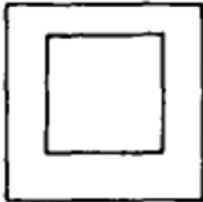
the British Ability Scales document, page 12-26 on <http://www.cls.ioe.ac.uk/page.aspx?&siteid=807&siteSectionTitle=Questionnaires>

The child has to complete each pattern by drawing the appropriate shape in the empty square.

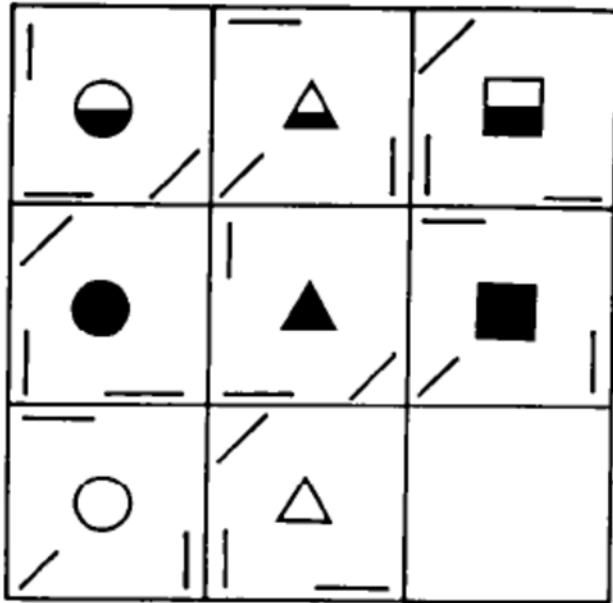
Example 1:



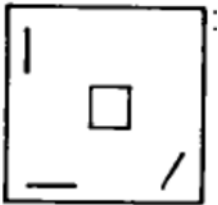
Solution:



Example 2:



Solution:



### 3B. Measures of personality

At age 10, various measures of personality were obtained, some were based on questions answered by the children and some by their teachers.

Personality traits include measures of self-esteem (16 items,  $\alpha=0.69$ ), locus of control (16 items,  $\alpha=0.63$ ). These are based on questions answered by the respondents. For a full list of items, see the Pupil Questionnaire Form on <http://www.cls.ioe.ac.uk/page.aspx?&siteid=807&sitecontentid=Questionnaires>. Self-esteem is measured by items k010-k025. Locus of control is measured by items k075-k094.

The data also contain personality traits answered by the pupils' teachers: 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,  $\alpha=0.58$ ). For a full list of

items, see

<http://www.cls.ioe.ac.uk/datadictionary/page.asp?section=000100010002000600040004&sectionTitle=2+-+C%3A+Child%27s+Developmental+Behaviours>. Disorganized activity is measured by items j127 j129 j138 j139 j143 j152 j155 j158 j162 j174 j177. Anti-social behavior is measured by items j131 j134 j135 j142 j147 j160 j163 j169 j170 j175. Neuroticism is measured by items j128 j145 j146 j149 j178b and Introversion is measured by items j148 j159 j162 j170 j178a.

### 3C. Achievement tests

The BCS includes several achievement tests, measured at age 10 and 16. Below we give a short description and some examples of the test we use.

#### 3C.1 The BAS achievement test

At age 10, the children take the BAS, which is an achievement test consisting of vocabulary (37 definitions of words), recalling digits (34 sequences), 21 tests of word similarities (the children here three words and need to mention what would fit to these words). See, for all items:

<http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=807&sitesectiontitle=Questionnaires>.

#### 3C.2 The Friendly Maths Test (FMT)

At age 10, the children took the Friendly Maths Test. This is a test consisting of 72 questions of the type indicated below. The test is taken at age 10. For all items, see:

<http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=807&sitesectiontitle=Questionnaires>

Example 1:

**What is the  
MISSING  
number ?**

$$\begin{array}{r} 5 \\ + ? \\ \hline 8 \end{array}$$

3   
13   
4   
5   
2

Example 2:

How many STARS altogether ?

7

4

11

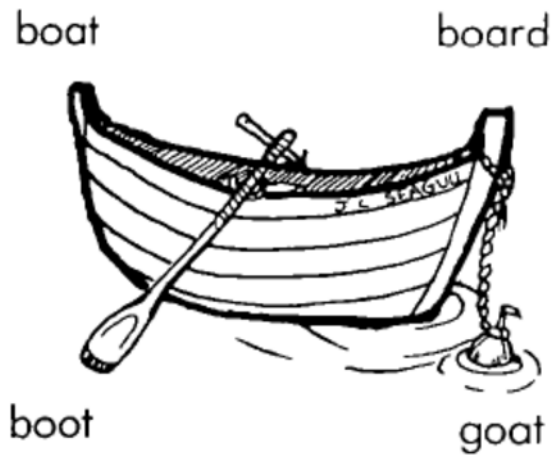
10

9

3C.3 The Shortened Edinburgh Reading Test (ERT)

The Shortened Edinburgh Reading Test consists of 54 questions of the type indicated below. The test is taken at age 10. For all items, see: <http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=807&sitesectiontitle=Questionnaires>

Example 1:



Example 2:

**Read this passage, then decide how each item should be completed to reproduce the sense of the passage. Underline the best completion**

I locked the door, knowing it would hold my pursuers for a few moments, and attempted to open the window. It would not budge. Sheer blind panic seized my mind as the panels splintered, but then my fingers found the catch, the window opened, and I was breathing the cold air. There was a deafening blast, and something struck my shoulder, pushing me off the ledge.

Had there not been a large rose bush below I would not have got up, for my fall was all arms and legs, naturally enough. As it was, I arose scratched and furious. My panic had gone completely, and I wished only to avenge myself on my attackers, for the idea of flight, which had previously obsessed me, had now left me entirely. I stood there with the snow melting on my knees, shaking my fist at my pursuers, who fortunately could not see me in the dark.

1 This passage describes (a kidnapping / a man's escape from attackers /  
/ police raiding a house / a murder) I3063

### *3C.4 The Ches Pictorial Language Comprehension Test (PLCT)*

The test is taken at age 10. The children see a word or a sentence and have to choose the right picture which matches the word/sentence, or place pictures in the correct order so that it matches the sentence. There are 103 questions to be answered in total. See <http://www.cls.ioe.ac.uk/page.aspx?&siteid=807&siteidtitle=Questionnaires>

### *3C.5 A vocabulary test at age 16*

The test is taken at age 10. The children get a list of 100 words and need to indicate if they are spelled correctly. See page 42 in Document B and the last two pages of Document C on:

<http://www.cls.ioe.ac.uk/page.aspx?&siteid=832&siteidtitle=Questionnaires>

### *3C.6 A spelling test at age 16*

The children get a word and need to choose a similar word from five alternatives. There are 75 questions to be answered. For all items see page 27-29 in Document B on: <http://www.cls.ioe.ac.uk/page.aspx?&siteid=832&siteidtitle=Questionnaires>

### *3C7. The scores on a Math test at age 16*

The children answer 60 arithmetic test questions. For all items see page 23-26 in Document B on:

<http://www.cls.ioe.ac.uk/page.aspx?&siteid=832&sitecontentid=Questionnaires>

### 3D. Grades

We have information on the grades of 14 high school subjects at age 16. These include English Language, English literature, Mathematics, Science, Physics, Biology, History, Geography, Chemistry, French, German, Business Communication, RE (Religious Education), and Home Economics. More information about this question can be found in Document T (Family Follow-Up Form), page 3 on:

<http://www.cls.ioe.ac.uk/page.aspx?&siteid=832&sitecontentid=Questionnaires>

### 3E. Life outcomes

We use: 1. Wages (at age 38), 2. Educational attainment (we use the nominal age at which the highest degree is obtained, measured at age 34), 3. The Body Mass Index (age 34), 4. Number of times been arrested and taken to a police station (age 34), 5. Satisfaction with life so far (age 34).

## Appendix 4: Description of the National Longitudinal Study of Youth, 1979 (NLSY)

The NLSY79 has Armed Forces Qualifying Test (AFQT) scores, scores on IQ tests, two measures of personality---the Rotter measure of locus of control and Rosenberg measure of self-esteem which are also included as *z*-scores. Results from the NLSY79 should be carefully interpreted with the understanding that the two personality measures are quite limited, and the IQ information is based on different IQ-tests. 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 *z*-scores constructed from IQ percentiles, which---in theory---should be comparable across tests, allowing us to pool test scores from the IQ tests for a much larger sample of test takers.

### 4A. Non-cognitive Measurements

#### *Rosenberg Self Esteem Scale*

The Rosenberg Self Esteem Scale was administered in 1979 and they ask:

1. I feel that I am a person of worth, at least on an equal plane with others.
2. I feel that I have a number of good qualities.
3. All in all, I am inclined to feel that I am a failure.
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of.
6. I take a positive attitude toward myself.
7. On the whole, I am satisfied with myself.
8. I wish I could have more respect for myself.
9. I certainly feel useless at times.
10. At times I think I am no good at all.

#### *Rotter Locus of Control*

Four Questions from the Rotter Locus of Control were administered in 1979. The rotter locus of control asks the individual to choose which of two statements they most agree with, then to specify if they slightly agree or strongly agree. The four questions:

1. What happens to me is my own doing / Sometimes I feel that I don't have enough control over the direction my life is taking.
2. When I make plans, I am almost that I can make them work / It is not always wise to plan too far ahead, because many things turn out to be a matter of good or bad fortune anyhow.

3. In my case, getting what I want has little or nothing to do with luck / Many times we might just as well decide what to do by flipping a coin.
4. Many times I feel that I have little influence over the things that happen to me / It is impossible for me to believe that chance or luck plays an important role in my life.

*IQ tests*

IQ test scores were recorded in the NLSY for Individual's high school transcripts who signed the transcript release waivers. Year, grade level, age, and percentile score for these tests are also reported. The recorded tests are:

- CTMM - California Test of Mental Maturity
- LTIT - Lorge-Thorndike Intelligence Test
- HNTMM - Henmon-Nelson Test of Mental Maturity
- KAIT - Kuhlmann-Anderson Intelligence Test
- SBIS - Stanford-Binet Intelligence Scale
- WCIS - Wechsler Intelligence Scale for Children
- DAT - Differential Aptitude Test
- CSCAT - Coop School and College Ability Test

There are a few concerns we face with these IQ tests. First off, these tests were administered from the late sixties to the late seventies making them older IQ tests, some of which have fallen out of favor or have been updated to eliminate cultural or language dependencies.

The Differential Aptitude Test and the Coop School and College Ability Test have been excluded from this list as they are achievement tests, not IQ tests.

**CTMM**

The CTMM is a group intelligence test that was discontinued in the 1980s. In *Measurement and Evaluation in Psychology and Education*, Thorndike and Hagen (1977) list the 1963 edition of the test, so it is highly likely that this was the version used in 1979. The 1963 revision of the test could be administered in either short or full form, and was differentiated into 8 levels of tests, appropriate for different age levels. The test returned verbal and non-verbal IQ in addition to the standard total IQ score. The subtests composing each battery are shown below:

CTMM	CTMM - SF
Logical reasoning	Logical reasoning
Numerical ability	Numerical ability (abbreviated)
Verbal ability	Verbal ability (abbreviated)
Memory	Memory (abbreviated)
Spatial relations	

Question types in each subtest:



Logical Reasoning: Similarities, opposites, analogies  
 Spatial Relations: Lefts and rights, manipulation of areas  
 Numerical Reasoning: Number problems, numerical series, numerical values  
 Verbal Concepts: Verbal comprehension, inferences  
 Memory: Immediate recall, delayed recall

One of the possible complications in this test structure is the verbal nature of the most questions. This is confirmed in the factor analysis of Anderson and Leton, 1964; analogies (logic subtest) load significantly on verbal concepts, so do number problems and number series (numerical reasoning subtest).

**LTIT**

The description of the LTIT dates back to 1964. Of the five levels, levels 3 to 5 are verbal in nature, but parallel nonverbal tests are available. The latter consist of strictly pictorial, diagrammatical, and numerical questions. However, according to Siegel, tests are based on the premise that “abstract intelligence requires the manipulation of verbal symbols.” Hence, perhaps the loading of non-verbal questions on verbal skills may not be a severe problem.

The Cognitive Abilities Test (CogAT) descended from the LTIT and has largely supplanted it. For the bulk of children and young adults, the test consists of a verbal, quantitative and nonverbal section. The breakdown of subtests in these sections is as follows:

Verbal	Nonverbal	Quantitative
Vocabulary	Figure Analogies	Quantitative Comparison
Sentence Completion	Figure Classification	Number Series
Verbal Classification	Figure Synthesis	Equation Building
Verbal Analogies		

For the very young (primary grades), a 4 subtest version was created, consisting of Oral Vocabulary, Relational Concepts, Multimental, and Quantitative Concepts. This test relies on verbal instructions and is entirely pictorial on paper.

**KAIT**

- Scores are comparable with Stanford-Binet
- Returns Mental Age, IQ, and Percent of Average Development (preferred)

**DAT**

- 8 subtests: Verbal Reasoning, Abstract Reasoning, Numerical Reasoning, Perceptual Speed and Accuracy, Mechanical Reasoning, Space Relations, Spelling, Language Usage

- VR + SR measures school ability, hence removal of SA score may give unmeasured (by AFQT) cognitive ability

### **Stanford-Binet**

- Widely used
- Verbally loaded, in the 1979 version, and returns a single factor score

### **WISC**

- 10 subtests (1979): Information, similarities, vocabulary, block design, picture arrangement, arithmetic, coding, object assembly, picture completion, comprehension
- Research has shown that low school performance strongly correlated with verbal cluster skills (vocabulary, information, similarities, comprehension)

## Appendix 5: Description of the National Survey of Midlife Development in the U.S. (MIDUS)

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 are comprised of four parts (<http://aging.wisc.edu/midus/midus1/index.php>). 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 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 first round of MIDUS included a number of “studies within a study” such as diary studies of daily stress carried out on a subsample of the respondents. (see <http://midus.wisc.edu/scopeofstudy.php> for details).

Based on the success of MIDUS a second round of interviews provides follow-up data of the same individuals 7 to 11 years later (MIDUS II). Similar to MIDUS, MIDUS II included a number of sub-studies including information collected on health behaviors, psychological and social factors, health and illness, and neurobiological mechanisms such as cortisol levels, blood pressure, and indicators of inflammation.

In both MIDUS and MIDUS II, Big Five personality traits were measured by 30 questions on how well the respondent is described by descriptive adjectives. Given a word, respondent choose how well the words describes them from (1) A lot, (2) some, (3) a little), and (4) not at all (see Rossi (2001) for details). The scale measures the Big Five personality traits: Neuroticism (alpha = 0.74), Extraversion (alpha = 0.78), Openness (alpha = 0.77), Conscientiousness (alpha = 0.58), and agreeableness (alpha = 0.80). Additional information on the scales as well as the questionnaires used can be found at <http://midus.wisc.edu/midus2/project1/>.

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 (2006) and Lachman and Tun (2008) for details. More information about the BTACT can be found at <http://midus.wisc.edu/midus2/project3/> and [www.brandeis.edu/projects/lifespan](http://www.brandeis.edu/projects/lifespan).

The analysis in this paper focuses on those aged 30 to 60 during MIDUS II who were part of the main sample and have non-missing cognitive and Big Five measures leaving us with a main sample of 2,289 observations. While MIDUS includes measures of

personality, cognitive ability, and outcomes, it does not provide scores on achievement tests.

## Appendix 6: Pairwise Pearson Correlations

The pairwise correlations among grades, IQ, and achievement are strong, but far from perfect.

Table 6.1: Correlations (Pearson Correlations)

<b>All</b>	Stella Maris	BCS	NLSY	MIDUS
$\rho$ (IQ, Achievement)	0.378 (0.000)	0.509 (0.000)	0.698 (0.000)	
$\rho$ (IQ, Grades)	0.112 (0.054)	0.338 (0.000)	0.464 (0.000)	
$\rho$ (IQ, Personality)	0.195 (0.048)	0.451 (0.000)	0.291 (0.000)	0.189 (0.000)
$\rho$ (Achievement, Grades)	0.316 (0.000)	0.379 (0.000)	0.61 (0.000)	
$\rho$ (Achievement, Personality)	0.294 (0.000)	0.446 (0.000)	0.41 (0.000)	
$\rho$ (Grades, Personality)	0.257 (0.003)	0.433 (0.000)	0.305 (0.000)	

Note:  $p$ -values reported in parentheses.

Table 6.2: Correlations Stella Maris data – Full data set

	GPA	DAT	Raven	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	Grit
GPA	1								
DAT	0.3164 0.0000	1							
Raven	0.1115 0.0544	0.3783 0.0000	1						
Openness	0.0517 0.3719	0.2204 0.0001	0.0995 0.0648	1					
Conscientiousness	0.2000 0.0005	-0.0123 0.8261	0.1000 0.0635	0.1223 0.0227	1				
Extraversion	-0.0414 0.4747	-0.0872 0.1182	-0.0714 0.1859	0.2380 0.0000	0.0272 0.6140	1			
Agreeableness	0.0256 0.6583	0.0046 0.9338	0.0153 0.7767	0.2669 0.0000	0.1600 0.0028	0.2861 0.0000	1		
Neuroticism	-0.0563 0.3310	-0.0546 0.3288	-0.0834 0.1220	0.0441 0.4128	-0.1595 0.0029	-0.1384 0.0098	-0.0342 0.5257	1	
Grit	0.2350 0.0001	-0.0094 0.8690	0.0682 0.2149	0.3121 0.0000	0.6140 0.0000	0.1131 0.0389	0.2106 0.0001	-0.0728 0.1842	1

Notes: Stella Maris data. This table shows the correlations one by one between a variable at the top of the column and a variable in a row. The first number is the correlation coefficient, the number below it is the *p*-value related to the correlation.

Table 6.3: Correlations BCS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. IQ	1																						
2. Self-Esteem	0.156 0.000	1																					
3. LoC	0.352 0.000	0.449 0.000	1																				
4. Disorganized	-0.400 0.000	-0.220 0.000	-0.401 0.000	1																			
5. Anti-Social	-0.205 0.000	-0.182 0.000	-0.192 0.000	0.557 0.000	1																		
6. Neuroticism	-0.165 0.000	-0.157 0.000	-0.203 0.000	0.420 0.000	0.256 0.000	1																	
7. Introversion	-0.148 0.000	-0.153 0.000	-0.207 0.000	0.450 0.000	0.309 0.000	0.548 0.000	1																
8. BAS Sim	0.485 0.000	0.172 0.000	0.398 0.000	-0.331 0.000	-0.136 0.000	-0.156 0.000	-0.143 0.000	1															
9. BAS Words	0.448 0.000	0.192 0.000	0.415 0.000	-0.334 0.000	-0.132 0.000	-0.163 0.000	-0.132 0.000	0.623 0.000	1														
10. BAS Digits	0.298 0.000	0.118 0.000	0.270 0.000	-0.260 0.000	-0.102 0.000	-0.141 0.000	-0.114 0.000	0.320 0.000	0.312 0.000	1													
11. BAS	0.532 0.000	0.208 0.000	0.467 0.000	-0.388 0.000	-0.156 0.000	-0.189 0.000	-0.163 0.000	0.888 0.000	0.884 0.000	0.515 0.000	1												
12. PCLT	0.509 0.000	0.184 0.000	0.394 0.000	-0.331 0.000	-0.149 0.000	-0.162 0.000	-0.138 0.000	0.583 0.000	0.659 0.000	0.307 0.000	0.691 0.000	1											
13. FMT	0.641 0.000	0.229 0.000	0.478 0.000	-0.498 0.000	-0.204 0.000	-0.237 0.000	-0.210 0.000	0.594 0.000	0.579 0.000	0.390 0.000	0.672 0.000	0.593 0.000	1										
14. ERT	0.588 0.000	0.204 0.000	0.482 0.000	-0.510 0.000	-0.228 0.000	-0.213 0.000	-0.198 0.000	0.585 0.000	0.602 0.000	0.379 0.000	0.677 0.000	0.612 0.000	0.733 0.000	1									
15. Spelling	0.316 0.000	0.148 0.000	0.287 0.000	-0.313 0.000	-0.153 0.000	-0.098 0.000	-0.075 0.000	0.365 0.000	0.390 0.000	0.187 0.000	0.423 0.000	0.376 0.000	0.400 0.000	0.439 0.000	1								
16. Vocabulary	0.152 0.000	0.079 0.000	0.128 0.000	-0.200 0.000	-0.115 0.000	-0.040 0.005	-0.073 0.000	0.137 0.000	0.130 0.000	0.084 0.000	0.151 0.000	0.126 0.000	0.176 0.000	0.208 0.000	0.470 0.000	1							
17. Math	0.513 0.000	0.190 0.000	0.404 0.000	-0.473 0.000	-0.218 0.000	-0.208 0.000	-0.191 0.000	0.429 0.000	0.438 0.000	0.321 0.000	0.507 0.000	0.421 0.000	0.666 0.000	0.575 0.000	0.604 0.000	0.305 0.000	1						
18. Grades	0.338 0.000	0.158 0.000	0.347 0.000	-0.355 0.000	-0.168 0.000	-0.159 0.000	-0.085 0.000	0.364 0.000	0.407 0.000	0.210 0.000	0.438 0.000	0.379 0.000	0.464 0.000	0.445 0.000	0.319 0.000	0.135 0.000	0.470 0.000	1					
19. Wage	0.153 0.000	0.146 0.000	0.201 0.000	-0.105 0.000	-0.019 0.209	-0.135 0.000	-0.084 0.000	0.228 0.000	0.282 0.000	0.089 0.000	0.282 0.000	0.231 0.000	0.278 0.000	0.207 0.000	0.165 0.000	0.049 0.010	0.247 0.000	0.243 0.000	1				
20. Education	0.339 0.000	0.154 0.000	0.307 0.000	-0.316 0.000	-0.155 0.000	-0.118 0.000	-0.090 0.000	0.339 0.000	0.387 0.000	0.194 0.000	0.410 0.000	0.346 0.000	0.417 0.000	0.403 0.000	0.326 0.000	0.156 0.000	0.413 0.000	0.438 0.000	0.329 0.000	1			
21. BMI	-0.089 0.000	-0.062 0.000	-0.100 0.000	0.130 0.000	0.114 0.000	-0.026 0.027	0.056 0.000	-0.052 0.000	-0.037 0.000	-0.058 0.000	-0.056 0.000	-0.038 0.001	-0.085 0.000	-0.116 0.000	-0.076 0.000	-0.061 0.000	-0.069 0.001	-0.111 0.000	0.024 0.090	-0.129 0.000	1		
22. Arrests	-0.064 0.000	-0.025 0.036	-0.036 0.002	0.092 0.000	0.094 0.000	0.002 0.835	0.027 0.021	-0.051 0.000	-0.043 0.000	-0.002 0.876	-0.048 0.000	-0.040 0.001	-0.043 0.000	-0.054 0.000	-0.050 0.001	-0.033 0.032	-0.065 0.001	-0.043 0.004	0.025 0.069	-0.089 0.000	0.012 0.267	1	
23. Life Satisfaction	0.081 0.000	0.094 0.000	0.105 0.000	-0.117 0.000	-0.074 0.000	-0.081 0.000	-0.117 0.000	0.046 0.000	0.050 0.000	0.027 0.023	0.052 0.000	0.058 0.000	0.092 0.000	0.083 0.000	0.053 0.001	0.049 0.001	0.112 0.000	0.086 0.000	0.052 0.000	0.102 0.000	-0.031 0.003	-0.112 0.000	1

Notes: BCS data. This table shows the correlations one by one between a variable at the top of the column and a variable in a row. The first number is the correlation coefficient, the number below it is the  $p$ -value related to the correlation.

Table 6.4: MIDUS Pairwise Pearson Correlations

	IQ	Consc.	Extrav	Neurot.	Agree.	Open.
IQ	1					
Consc.	0.0724 0.0004	1				
Extrav.	0.0395 0.056	0.25 0.0000	1			
Neurot.	-0.1051 0.000	-0.1719 0.000	-0.1472 0.000	1		
Agree.	-0.0315 0.127	0.2634 0.0000	0.4895 0.0000	-0.0284 0.1262	1	
Open.	0.1425 0.0000	0.2487 0.0000	0.5039 0.0000	-0.1795 0.0000	0.3284 0.0000	1

Note: This Table shows pairwise Pearson correlations between fluid intelligence and Big Five personality traits in the MIDUS data set. *p*-values are reported below each correlation.

Table 6.5: NLYS79 Pairwise Pearson Correlations

	IQ	Rotter	Rosenberg	GPA	AFQT
IQ	1				v
Rotter	0.2215 0.0000	1			
Rosenberg	0.2225 0.0000	0.2779 0.0000	1		
GPA	0.4585 0.0000	0.1727 0.0000	0.2057 0.0000	1	
AFQT	0.7041 0.0000	0.3187 0.0000	0.3361 0.0000	0.5889 0.0000	1

Note: This Table shows the pairwise Pearson correlation between IQ, Rotter locus of control, Rosenberg self-esteem, 9th grade core subject GPA, and AFQT achievement test scores. *p*-values are reported below each correlation.

## Appendix 7: Decomposing Achievement Tests and Grades into IQ and Personality

Table 7.1: Decomposing Grades into IQ and Personality – Stella Maris

	(1) Grades	(2) Grades	(3) Grades	(4) Grades	(5) Grades
IQ	0.112* (0.058)			0.081 (0.059)	0.084 (0.059)
Openness		0.043 (0.063)		0.028 (0.064)	-0.022 (0.067)
Conscientiousness		0.185*** (0.058)		0.179*** (0.059)	0.060 (0.075)
Extraversion		-0.056 (0.062)		-0.048 (0.063)	-0.059 (0.063)
Agreeableness		-0.004 (0.062)		0.004 (0.063)	-0.000 (0.063)
Neuroticism		-0.040 (0.058)		-0.030 (0.058)	-0.033 (0.058)
Grit			0.233*** (0.057)		0.202** (0.078)
Constant	-0.001 (0.058)	-0.000 (0.057)	0.006 (0.057)	-0.002 (0.057)	0.001 (0.058)
Observations	298	300	291	298	289
Adjusted $R^2$	0.009	0.028	0.052	0.031	0.050

Notes: The Stella Maris data include 347 Dutch high school students aged 15 or 16 in 2008. The table shows regressions of Grades on (1) IQ, (2) Grades on the Big Five, (3) Grades on Grit, (4) Grades on IQ and the Big Five, (5) 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. Grades are 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. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



Table 7.2: Decomposing DAT Achievement Test into IQ and Personality – Stella Maris

	(1) DAT	(2) DAT	(3) DAT	(4) DAT	(5) DAT
IQ	0.376*** (0.052)			0.346*** (0.051)	0.341*** (0.052)
Openness		0.279*** (0.058)		0.232*** (0.055)	0.242*** (0.058)
Conscientiousness		-0.061 (0.055)		-0.078 (0.052)	-0.014 (0.067)
Extraversion		-0.163*** (0.059)		-0.124** (0.056)	-0.134** (0.056)
Agreeableness		-0.018 (0.058)		-0.015 (0.055)	0.001 (0.056)
Neuroticism		-0.105* (0.056)		-0.075 (0.053)	-0.065 (0.053)
Grit			-0.009 (0.056)		-0.085 (0.069)
Constant	-0.007 (0.052)	-0.002 (0.054)	-0.007 (0.056)	-0.006 (0.051)	-0.013 (0.051)
Observations	320	322	311	320	309
Adjusted $R^2$	0.140	0.067	-0.003	0.181	0.180

Notes: The Stella Maris data include 347 Dutch high school students aged 15 or 16 in 2008. The table shows regressions of DAT on (1) IQ, (2) DAT on the Big Five, (3) DAT on Grit, (4) DAT on IQ and the Big Five, (5) DAT 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. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 7.3: Decomposing Achievement Tests into IQ and Personality – BCS

	(1) BAS Similarities	(2) BAS Similarities	(3) BAS Similarities	(4) BAS Word definitions	(5) BAS Word definitions	(6) BAS Word definitions
IQ	0.491*** (0.010)		0.372*** (0.010)	0.449*** (0.010)		0.316*** (0.010)
Self-esteem		-0.015 (0.011)	-0.011 (0.010)		0.005 (0.011)	0.009 (0.010)
Locus of control		0.325*** (0.012)	0.239*** (0.011)		0.333*** (0.012)	0.260*** (0.011)
Disorganized		-0.234*** (0.014)	-0.109*** (0.013)		-0.232*** (0.013)	-0.127*** (0.013)
Anti-social		0.051*** (0.012)	0.047*** (0.011)		0.054*** (0.012)	0.051*** (0.011)
Neuroticism		-0.021* (0.012)	-0.018 (0.011)		-0.034*** (0.012)	-0.031*** (0.011)
Introversion		0.021* (0.012)	0.003 (0.011)		0.042*** (0.012)	0.027** (0.012)
Constant	-0.000 (0.009)	0.003 (0.010)	-0.002 (0.009)	-0.004 (0.010)	-0.001 (0.010)	-0.006 (0.009)
Observations	8,431	8,431	8,431	8,501	8,501	8,501
Adjusted $R^2$	0.240	0.198	0.308	0.204	0.210	0.290

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. Scores on a Math test. Grades is the average grade of 14 subjects at age 16. The table shows 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. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 7.4: Decomposing Achievement Tests into IQ and Personality – BCS (continued)

	(1) BAS Digit recall	(2) BAS Digit recall	(3) BAS Digit recall	(4) BAS Achievement	(5) BAS Achievement	(6) BAS Achievement
IQ	0.289*** (0.010)		0.189*** (0.011)	0.435*** (0.007)		0.315*** (0.008)
Self-esteem		-0.011 (0.012)	-0.009 (0.011)		-0.006 (0.009)	-0.002 (0.008)
Locus of control		0.196*** (0.012)	0.152*** (0.012)		0.303*** (0.009)	0.230*** (0.009)
Disorganized		-0.200*** (0.014)	-0.137*** (0.015)		-0.225*** (0.011)	-0.119*** (0.010)
Anti-social		0.050*** (0.013)	0.048*** (0.012)		0.052*** (0.009)	0.049*** (0.008)
Neuroticism		-0.044*** (0.013)	-0.042*** (0.012)		-0.029*** (0.009)	-0.027*** (0.009)
Introversion		0.031** (0.013)	0.022* (0.013)		0.031*** (0.010)	0.016* (0.009)
Constant	0.005 (0.010)	0.007 (0.010)	0.004 (0.010)	-0.001 (0.007)	0.001 (0.008)	-0.003 (0.007)
Observations	8,497	8,497	8,497	8,408	8,408	8,408
Adjusted $R^2$	0.084	0.099	0.127	0.287	0.270	0.390

Notes: See Table 7.3.

Table 7.5: Decomposing Achievement Tests into IQ and Personality – BCS (continued)

	(1) PCLT	(2) PCLT	(3) PCLT	(4) FMT	(5) FMT	(6) FMT	(7) ERT	(8) ERT	(9) ERT
IQ	0.514*** (0.010)		0.402*** (0.010)	0.629*** (0.008)		0.466*** (0.008)	0.584*** (0.009)		0.400*** (0.009)
Self-esteem		0.001 (0.011)	0.005 (0.010)		0.014 (0.010)	0.020** (0.008)		-0.017* (0.010)	-0.012 (0.009)
Locus of control		0.319*** (0.012)	0.230*** (0.011)		0.311*** (0.010)	0.203*** (0.009)		0.333*** (0.010)	0.240*** (0.010)
Disorganized		-0.225*** (0.014)	-0.093*** (0.013)		-0.411*** (0.012)	-0.256*** (0.011)		-0.428*** (0.012)	-0.296*** (0.011)
Anti-social		0.024* (0.012)	0.021* (0.011)		0.083*** (0.011)	0.079*** (0.009)		0.054*** (0.011)	0.051*** (0.010)
Neuroticism		-0.022* (0.012)	-0.020* (0.011)		-0.039*** (0.011)	-0.035*** (0.009)		-0.005 (0.011)	-0.001 (0.010)
Introversion		0.040*** (0.013)	0.021* (0.012)		0.038*** (0.011)	0.016* (0.009)		0.051*** (0.011)	0.033*** (0.010)
Constant	-0.017* (0.010)	-0.005 (0.010)	-0.018* (0.009)	0.019** (0.008)	0.022** (0.009)	0.015** (0.007)	0.012 (0.009)	0.013 (0.009)	0.007 (0.008)
Observations	7,960	7,960	7,960	8,486	8,486	8,486	8,506	8,506	8,506
Adjusted $R^2$	0.264	0.199	0.328	0.409	0.344	0.523	0.344	0.352	0.481

Notes: See Table 7.3.

Table 7.6: Decomposing Achievement Tests into IQ and Personality – BCS (continued)

	(1) Spelling	(2) Spelling	(3) Spelling	(4) Vocabulary	(5) Vocabulary	(6) Vocabulary	(7) Math	(8) Math	(9) Math
IQ	0.327*** (0.017)		0.201*** (0.018)	0.152*** (0.017)		0.072*** (0.019)	0.516*** (0.019)		0.348*** (0.019)
Self-esteem		0.031* (0.018)	0.030* (0.017)		0.023 (0.019)	0.023 (0.019)		-0.002 (0.021)	-0.006 (0.019)
Locus of control		0.186*** (0.020)	0.142*** (0.020)		0.067*** (0.021)	0.051** (0.021)		0.275*** (0.023)	0.199*** (0.022)
Disorganized		-0.302*** (0.023)	-0.236*** (0.024)		-0.205*** (0.025)	-0.181*** (0.025)		-0.420*** (0.027)	-0.316*** (0.025)
Anti-social		-0.001 (0.022)	-0.005 (0.022)		-0.041* (0.024)	-0.043* (0.023)		0.032 (0.026)	0.031 (0.024)
Neuroticism		0.024 (0.020)	0.026 (0.019)		0.076*** (0.021)	0.076*** (0.021)		-0.001 (0.023)	0.007 (0.022)
Introversion		0.075*** (0.020)	0.070*** (0.020)		0.011 (0.022)	0.009 (0.022)		0.051** (0.024)	0.038* (0.022)
Constant	-0.052*** (0.017)	-0.071*** (0.016)	-0.096*** (0.016)	-0.019 (0.017)	-0.046*** (0.017)	-0.055*** (0.018)	-0.072*** (0.019)	-0.082*** (0.019)	-0.119*** (0.018)
Observations	3,341	3,341	3,341	3,341	3,341	3,341	2,033	2,033	2,033
Adjusted R <sup>2</sup>	0.104	0.141	0.173	0.022	0.051	0.055	0.260	0.288	0.385

Notes: See Table 7.3.

Table 7.7: Decomposing Grades into IQ and Personality – BCS (continued)

	(1) Grades	(2) Grades	(3) Grades
IQ	0.367*** (0.017)		0.220*** (0.017)
Self-esteem		-0.015 (0.018)	-0.019 (0.017)
Locus of control		0.267*** (0.019)	0.223*** (0.019)
Disorganized		-0.337*** (0.023)	-0.277*** (0.023)
Anti-social		0.002 (0.022)	-0.006 (0.021)
Neuroticism		-0.030 (0.019)	-0.023 (0.019)
Introversion		0.091*** (0.020)	0.081*** (0.019)
Constant	-0.103*** (0.017)	-0.143*** (0.016)	-0.186*** (0.016)
Observations	3,521	3,521	3,521
Adjusted $R^2$	0.114	0.189	0.223

Notes: See Table 7.3.

Table 7.8: Decomposing Achievement Tests and Grades into IQ and Personality – NLSY

	(1)	(2)	(3)	(4)	(5)	(6)
	AFQT	AFQT	AFQT	Grades	Grades	Grades
IQ	0.666*** (0.023)		0.601*** (0.023)	0.392*** (0.026)		0.347*** (0.027)
Rotter		0.284*** (0.033)	0.169*** (0.025)		0.187*** (-0.031)	0.125*** (0.029)
Rosenberg		0.255*** (0.032)	0.132*** (0.024)		0.158*** (0.030)	0.085*** (0.028)
Constant	-0.169*** (0.024)	-0.186*** (0.030)	-0.161*** (0.023)	2.431*** (0.027)	2.434*** (0.029)	2.435*** (0.026)
Observations	877	877	877	823	823	823
Adjusted $R^2$	0.489	0.173	0.538	0.216	0.093	0.248

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 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. 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 table shows 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.

## Appendix 8: Decomposing the Contributions of IQ and Personality to Life Outcomes

Table 8.1: Log wage income, Age 40 on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	0.173*** (0.040)		0.144*** (0.042)		0.016 (0.055)		0.098** (0.047)	0.016 (0.057)
Rotter		0.051 (0.045)	0.028 (0.045)		-0.0003 (0.045)		0.032 (0.046)	0.017 (0.046)
Rosenberg		0.138** (0.044)	0.108** (0.045)		0.080* (0.045)		0.078* (0.046)	0.062 (0.046)
AFQT				0.251*** (0.042)	0.214*** (0.061)			0.171* (0.068)
GPA						0.229*** (0.050)	0.146** (0.057)	0.087 (0.062)
Constant	10.215*** (0.041)	10.214*** (0.041)	10.214*** (0.041)	10.246*** (0.041)	10.240*** (0.041)	9.651*** (0.132)	9.852*** (0.148)	10.018*** (0.161)
Observations	554	554	554	554	554	529	529	529
Adjusted $R^2$	0.031	0.021	0.040	0.059	0.036	0.059	0.059	0.050

Source: NLSY. Each column reports the result of a regression analysis with log wage income as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



Table 8.2: SF12 Mental Health on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	41.505 (29.770)		10.845 (30.824)		-1.187 (40.988)		7.015 (35.074)	-1.366 (42.273)
Rotter		42.209 (32.689)	40.292 (33.162)		37.463 (33.785)		22.551 (34.197)	20.794 (34.576)
Rosenberg		104.280** (32.837)	101.959*** (33.516)		99.091*** (34.148)		102.047*** (34.462)	100.147*** (34.898)
AFQT				62.078** (31.375)	20.153 (45.214)			18.024 (50.667)
GPA						45.284 (36.822)	9.109 (41.985)	2.548 (45.886)
Constant	5,337.642** (30.795)	5,334.656*** (30.431)	5,335.614*** (30.573)	5,347.551*** (31.407)	5,338.866*** (31.451)	5,232.071*** (95.335)	5,319.263*** (107.895)	5,337.942*** (120.064)
Observations	643	643	643	643	643	604	604	604
Adjusted $R^2$	0.001	0.019	0.018	0.005	0.001	0.013	0.017	0.014

Source: NLSY. Each column reports the result of a regression analysis with mental health as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.3: SF12 Physical Health on IQ, Rosenberg, Rotter, AFQT, and Grades - NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	146.699*** (30.289)		135.422*** (31.609)		70.721* (41.858)		109.942*** (34.325)	67.551 (41.258)
Rotter		18.702 (33.996)	-5.233 (34.006)		-20.450 (34.502)		-36.061 (33.467)	-44.948 (33.746)
Rosenberg		84.488** (34.150)	55.503 (34.369)		40.079 (34.873)		41.152 (33.726)	31.541 (34.060)
AFQT				166.655*** (31.878)	108.373** (46.174)			91.163* (49.450)
GPA						100.101*** (36.103)	37.589 (41.089)	4.404 (44.784)
Constant	5,220.546** (31.332)	5,206.778*** (31.648)	5,218.736*** (31.352)	5,243.823*** (31.910)	5,236.227*** (32.118)	4,992.510*** (93.474)	5,148.342*** (105.592)	5,242.815*** (117.181)
Observations	643	643	643	643	643	604	604	604
Adjusted $R^2$	0.034	0.009	0.035	0.039	0.011	0.031	0.042	0.028

Source: NLSY. Each column reports the result of a regression analysis with physical health as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.4: Voted in 2006 election on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	0.073*** (0.018)		0.043** (0.018)			-0.004 (0.025)	-0.010 (0.024)	0.017 (0.021)
Rotter		0.061*** (0.019)	0.053*** (0.020)			0.041** (0.020)	0.040** (0.020)	0.046** (0.020)
Rosenberg		0.089*** (0.020)	0.079*** (0.020)			0.068*** (0.020)	0.067*** (0.020)	0.073*** (0.020)
AFQT				0.114*** (0.018)		0.044 (0.030)	0.087*** (0.027)	
GPA					0.113*** (0.022)	0.056** (0.027)		0.071*** (0.025)
Constant	0.706*** (0.018)	0.702*** (0.018)	0.706*** (0.018)	0.725*** (0.019)	0.432*** (0.056)	0.579*** (0.070)	0.720*** (0.018)	0.535*** (0.063)
Observations	610	610	610	610	573	573	610	573
Adjusted $R^2$	0.026	0.059	0.066	0.058	0.044	0.081	0.080	0.079

Source: NLSY. Each column reports the result of a regression analysis with whether one voted in the 2006 election as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.5: Ever on Welfare between age 31 and 45 on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	-0.086*** (0.012)		-0.071*** (0.013)			-0.007 (0.017)	-0.007 (0.017)	-0.054*** (0.014)
Rotter		-0.049*** (0.014)	-0.036** (0.014)			-0.023 (0.014)	-0.017 (0.014)	-0.035** (0.014)
Rosenberg		-0.049*** (0.014)	-0.034** (0.014)			-0.019 (0.014)	-0.020 (0.014)	-0.029** (0.014)
AFQT				-0.124*** (0.013)		-0.102*** (0.021)	-0.107*** (0.019)	
GPA					-0.085*** (0.015)	-0.001 (0.019)		-0.038** (0.017)
Constant	0.176*** (0.013)	0.178*** (0.013)	0.174*** (0.013)	0.156*** (0.013)	0.381*** (0.039)	0.160*** (0.048)	0.157*** (0.013)	0.265*** (0.043)
Observations	877	877	877	877	823	823	877	823
Adjusted R <sup>2</sup>	0.052	0.035	0.067	0.099	0.037	0.095	0.100	0.070

Source: NLSY. Each column reports the result of a regression analysis with whether one was ever on welfare between age 31 and 45 as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.6: Depression on IQ, Rosenberg, Rotter, AFQT, and Grades – NLSY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ	-0.841*** (0.145)		-0.659*** (0.149)			-0.294 (0.201)	-0.316 (0.198)	-0.581*** (0.168)
Rotter		-0.424*** (0.160)	-0.311* (0.160)			-0.145 (0.163)	-0.231 (0.162)	-0.206 (0.162)
Rosenberg		-0.697*** (0.161)	-0.560*** (0.162)			-0.475*** (0.165)	-0.479*** (0.164)	-0.539*** (0.164)
AFQT				-1.026*** (0.152)		-0.617** (0.241)	-0.576*** (0.218)	
GPA					-0.632*** (0.179)	0.105 (0.218)		-0.120 (0.200)
Constant	2.868*** (0.150)	2.934*** (0.150)	2.879*** (0.148)	2.716*** (0.152)	4.418*** (0.463)	2.508*** (0.571)	2.784*** (0.152)	3.152*** (0.515)
Observations	638	638	638	638	601	601	638	601
Adjusted R <sup>2</sup>	0.049	0.047	0.074	0.065	0.019	0.073	0.082	0.065

Source: NLSY. Each column reports the result of a regression analysis with depression as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2B for a description of the variables. Standard errors in parentheses.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.7: Log wage on IQ, Personality and Education – MIDUS

	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	0.131*** (0.022)		0.102*** (0.023)		0.047** (0.023)
Conscientiousness		0.113*** (0.026)	0.109*** (0.026)		0.101*** (0.026)
Extraversion		0.002 (0.027)	0.004 (0.027)		0.022 (0.027)
Neuroticism		-0.074** (0.035)	-0.065* (0.035)		-0.055 (0.034)
Agreeableness		-0.172*** (0.023)	-0.164*** (0.023)		-0.157*** (0.023)
Openness		0.114*** (0.027)	0.100*** (0.027)		0.068** (0.027)
High School Graduate				0.169 (0.144)	0.126 (0.142)
Some College				0.279** (0.142)	0.199 (0.140)
College or more				0.614*** (0.140)	0.471*** (0.140)
Constant	10.479*** (0.022)	10.633*** (0.083)	10.602*** (0.084)	10.091*** (0.137)	10.284*** (0.153)
Observations	1651	1651	1651	1649	1649
Adjusted $R^2$	0.018	0.050	0.060	0.048	0.084

Source: MIDUS. Each column reports the result of a regression analysis with log wage as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2 for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.8: Current health on IQ, Personality and Education – MIDUS

	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	0.117*** (0.022)		0.087*** (0.021)		0.056** (0.022)
Conscientiousness		0.133*** (0.022)	0.127*** (0.022)		0.120*** (0.022)
Extraversion		0.081*** (0.026)	0.085*** (0.025)		0.092*** (0.025)
Neuroticism		-0.241*** (0.032)	-0.231*** (0.032)		-0.224*** (0.032)
Agreeableness		-0.023 (0.023)	-0.017 (0.023)		-0.012 (0.023)
Openness		0.009 (0.024)	-0.005 (0.024)		-0.022 (0.025)
High School Graduate				0.322** (0.127)	0.225* (0.121)
Some College				0.417*** (0.125)	0.264** (0.122)
College or more				0.575*** (0.122)	0.377*** (0.122)
Constant	0.031 (0.021)	0.585*** (0.072)	0.560*** (0.072)	-0.410*** (0.119)	0.251* (0.136)
Observations	2015	2015	2015	2013	2013
Adjusted $R^2$	0.015	0.072	0.080	0.022	0.086

Source: MIDUS. Each column reports the result of a regression analysis with current health as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2 for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.9: Physical health on IQ, Personality and Education – MIDUS

	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	0.216*** (0.020)		0.179*** (0.019)		0.123*** (0.020)
Conscientiousness		0.161*** (0.020)	0.153*** (0.019)		0.145*** (0.019)
Extraversion		0.056** (0.024)	0.060** (0.023)		0.073*** (0.023)
Neuroticism		-0.214*** (0.030)	-0.192*** (0.029)		-0.176*** (0.029)
Agreeableness		-0.062*** (0.021)	-0.047** (0.020)		-0.041** (0.020)
Openness		0.078*** (0.023)	0.049** (0.023)		0.019 (0.023)
High School Graduate				0.495*** (0.114)	0.359*** (0.112)
Some College				0.698*** (0.112)	0.493*** (0.111)
College or more				0.958*** (0.110)	0.656*** (0.112)
Constant	0.122*** (0.019)	0.611*** (0.069)	0.557*** (0.068)	-0.605*** (0.107)	0.011 (0.127)
Observations	2298	2298	2298	2296	2296
Adjusted $R^2$	0.051	0.083	0.117	0.066	0.141

Source: MIDUS. Each column reports the result of a regression analysis with physical health as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2 for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



Table 8.10: Mental health on IQ, Personality and Education – MIDUS

	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	0.164*** (0.022)		0.116*** (0.021)		0.083*** (0.022)
Conscientiousness		0.125*** (0.021)	0.119*** (0.021)		0.115*** (0.021)
Extraversion		0.113*** (0.025)	0.116*** (0.025)		0.123*** (0.025)
Neuroticism		-0.397*** (0.030)	-0.383*** (0.030)		-0.373*** (0.030)
Agreeableness		-0.061*** (0.022)	-0.051** (0.022)		-0.048** (0.022)
Openness		0.043* (0.024)	0.024 (0.024)		0.006 (0.024)
High School Graduate				0.375*** (0.124)	0.233** (0.116)
Some College				0.558*** (0.122)	0.334*** (0.115)
College or more				0.713*** (0.120)	0.409*** (0.116)
Constant	0.043** (0.020)	0.954*** (0.069)	0.919*** (0.069)	-0.511*** (0.116)	0.568*** (0.131)
Observations	2297	2297	2297	2295	2295
Adjusted $R^2$	0.027	0.131	0.144	0.031	0.151

Source: MIDUS. Each column reports the result of a regression analysis with mental health as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2 for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.11: Depression on IQ, Personality and Education – MIDUS

	(1)	(2)	(3)	(4)	(5)
Cognitive Ability	-0.083*** (0.023)		-0.057** (0.023)		-0.036 (0.024)
Conscientiousness		-0.048* (0.025)	-0.045* (0.025)		-0.041* (0.025)
Extraversion		-0.108*** (0.028)	-0.110*** (0.028)		-0.114*** (0.028)
Neuroticism		0.295*** (0.036)	0.288*** (0.036)		0.281*** (0.036)
Agreeableness		0.118*** (0.022)	0.114*** (0.022)		0.112*** (0.022)
Openness		0.042 (0.027)	0.051* (0.027)		0.061** (0.028)
High School Graduate				-0.409*** (0.157)	-0.348** (0.152)
Some College				-0.445*** (0.154)	-0.348** (0.150)
College or more				-0.532*** (0.151)	-0.399*** (0.150)
Constant	0.071*** (0.023)	-0.598*** (0.077)	-0.581*** (0.077)	0.523*** (0.148)	-0.213 (0.162)
Observations	2298	2298	2298	2296	2296
Adjusted $R^2$	0.005	0.049	0.051	0.009	0.055

Source: MIDUS. Each column reports the result of a regression analysis with depression as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figure 2 for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.12: Log wage regressions on Personality, IQ, Achievement, Grades, and Schooling - BCS

	(1) Wage	(2) Wage	(3) Wage	(4) Wage	(5) Wage	(6) Wage	(7) Wage	(8) Wage	(9) Wage	(10) Wage
IQ	0.164*** (0.018)		0.105*** (0.020)		0.042* (0.022)		0.045 (0.028)	-0.015 (0.030)		-0.022 (0.031)
Self-esteem		0.068*** (0.019)	0.066*** (0.019)		0.059*** (0.020)		0.087*** (0.027)	0.085*** (0.027)		0.078*** (0.028)
Locus of Control		0.156*** (0.021)	0.134*** (0.021)		0.093*** (0.022)		0.079*** (0.029)	0.039 (0.030)		0.026 (0.031)
Disorganized		-0.007 (0.024)	0.025 (0.025)		0.031 (0.026)		0.067* (0.035)	0.056 (0.036)		0.090** (0.038)
Anti-social		0.058** (0.023)	0.051** (0.023)		0.051** (0.024)		0.059* (0.033)	0.052 (0.034)		0.049 (0.035)
Neuroticism		-0.115*** (0.021)	-0.109*** (0.021)		-0.090*** (0.022)		-0.112*** (0.028)	-0.085*** (0.029)		-0.090*** (0.029)
Introversion		0.000 (0.022)	-0.010 (0.022)		-0.019 (0.023)		-0.031 (0.029)	-0.039 (0.030)		-0.040 (0.031)
PCLT				0.245*** (0.018)	0.180*** (0.021)			0.159*** (0.029)		0.126*** (0.030)
Grades						0.232*** (0.024)	0.191*** (0.027)	0.162*** (0.028)		0.072** (0.030)
Education									0.326*** (0.017)	0.263*** (0.026)
Constant	-0.034* (0.018)	-0.025 (0.017)	-0.040** (0.018)	-0.034* (0.018)	-0.042** (0.018)	0.055** (0.023)	0.034 (0.026)	0.028 (0.027)	-0.039** (0.017)	-0.031 (0.029)
Observations	3,252	3,252	3,252	3,033	3,033	1,778	1,778	1,660	2,931	1,507
Adj. $R^2$	0.024	0.053	0.061	0.055	0.079	0.049	0.081	0.091	0.109	0.146

Source: BCS. Each column reports the result of a regression analysis with the variable at the top of the column as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figures 1B and 2A for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.13: Regression of Body Mass Index (BMI) on Personality, IQ, Achievement, Grades, and Schooling - BCS

	(1) BMI	(2) BMI	(3) BMI	(4) BMI	(5) BMI	(6) BMI	(7) BMI	(8) BMI	(9) BMI	(10) BMI
IQ	-0.098*** (0.014)		-0.045*** (0.016)		-0.056*** (0.018)		-0.026 (0.023)	-0.031 (0.025)		-0.023 (0.025)
Self-esteem		-0.016 (0.016)	-0.016 (0.016)		-0.020 (0.016)		-0.028 (0.022)	-0.034 (0.022)		-0.035 (0.022)
Locus of Control		-0.046*** (0.017)	-0.036** (0.017)		-0.044** (0.018)		-0.018 (0.025)	-0.022 (0.026)		-0.016 (0.026)
Disorganized		0.128*** (0.020)	0.114*** (0.020)		0.121*** (0.021)		0.062** (0.030)	0.077** (0.031)		0.068** (0.031)
Anti-social		0.060*** (0.018)	0.062*** (0.018)		0.059*** (0.019)		0.085*** (0.028)	0.078*** (0.028)		0.079*** (0.028)
Neuroticism		-0.118*** (0.017)	-0.119*** (0.017)		-0.120*** (0.018)		-0.103*** (0.024)	-0.117*** (0.024)		-0.115*** (0.024)
Introversion		0.055*** (0.018)	0.056*** (0.018)		0.056*** (0.018)		0.043* (0.025)	0.046* (0.026)		0.048* (0.026)
PCLT				-0.052*** (0.015)	0.027 (0.017)			0.031 (0.025)		0.039 (0.025)
Grades						-0.103*** (0.019)	-0.066*** (0.021)	-0.071*** (0.023)		-0.047* (0.024)
Education									-0.135*** (0.014)	-0.072*** (0.022)
Constant	0.020 (0.014)	0.033** (0.014)	0.037*** (0.014)	0.005 (0.014)	0.032** (0.014)	-0.065*** (0.019)	-0.013 (0.021)	-0.026 (0.022)	0.005 (0.014)	-0.018 (0.022)
Observations	5,133	5,133	5,133	4,773	4,773	2,573	2,573	2,393	5,128	2,392
Adj. $R^2$	0.009	0.034	0.035	0.002	0.037	0.011	0.026	0.029	0.018	0.033

Source: BCS. Each column reports the result of a regression analysis with the variable at the top of the column as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figures 1B and 2A for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.14: Regression of Education on Personality, IQ, Achievement, and Grades – BCS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Education	Education	Education	Education	Education	Education	Education	Education
IQ	0.337*** (0.013)		0.209*** (0.014)		0.143*** (0.016)		0.139*** (0.021)	0.102*** (0.023)
Self-esteem		0.009 (0.014)	0.011 (0.014)		0.009 (0.014)		-0.008 (0.020)	-0.011 (0.021)
Locus of Control		0.219*** (0.016)	0.171*** (0.016)		0.129*** (0.016)		0.115*** (0.023)	0.089*** (0.024)
Disorganized		-0.271*** (0.018)	-0.204*** (0.018)		-0.194*** (0.019)		-0.120*** (0.028)	-0.126*** (0.029)
Anti-social		0.002 (0.017)	-0.008 (0.017)		-0.011 (0.017)		0.015 (0.026)	0.008 (0.026)
Neuroticism		0.017 (0.016)	0.023 (0.016)		0.034** (0.016)		0.009 (0.022)	0.024 (0.023)
Introversion		0.052*** (0.016)	0.046*** (0.016)		0.039** (0.016)		0.048** (0.023)	0.039 (0.024)
PCLT				0.352*** (0.014)	0.179*** (0.016)			0.128*** (0.023)
Grades						0.459*** (0.018)	0.350*** (0.020)	0.329*** (0.021)
Constant	-0.051*** (0.013)	-0.050*** (0.013)	-0.069*** (0.013)	-0.037*** (0.013)	-0.074*** (0.013)	0.233*** (0.018)	0.134*** (0.020)	0.107*** (0.021)
Observations	5,271	5,271	5,271	4,906	4,906	2,643	2,643	2,461
Adj. $R^2$	0.108	0.143	0.177	0.118	0.198	0.194	0.234	0.247

Source: BCS. Each column reports the result of a regression analysis with the variable at the top of the column as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figures 1B and 2A for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.15: Regressions of Arrests on Personality, IQ, Achievement, Grades, and Schooling – BCS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Arrests	Arrests	Arrests	Arrests	Arrests	Arrests	Arrests	Arrests	Arrests	Arrests
IQ	-0.071*** (0.014)		-0.048*** (0.015)		-0.042** (0.017)		-0.025 (0.018)	-0.007 (0.020)		-0.006 (0.020)
Self-esteem		0.009 (0.015)	0.009 (0.015)		0.005 (0.016)		0.011 (0.017)	0.007 (0.018)		0.007 (0.018)
Locus of Control		-0.003 (0.016)	0.008 (0.016)		0.013 (0.018)		0.036* (0.019)	0.032 (0.020)		0.032 (0.020)
Disorganized		0.084*** (0.019)	0.069*** (0.020)		0.074*** (0.021)		0.053** (0.024)	0.044* (0.024)		0.044* (0.025)
Anti-social		0.065*** (0.018)	0.068*** (0.018)		0.071*** (0.018)		0.015 (0.022)	0.022 (0.022)		0.022 (0.022)
Neuroticism		-0.051*** (0.017)	-0.052*** (0.016)		-0.065*** (0.017)		-0.032* (0.018)	-0.041** (0.019)		-0.041** (0.019)
Introversion		0.003 (0.017)	0.005 (0.017)		0.009 (0.018)		-0.002 (0.019)	0.001 (0.020)		0.001 (0.020)
PCLT				-0.049*** (0.014)	-0.013 (0.017)			-0.025 (0.019)		-0.024 (0.020)
Grades						-0.049*** (0.015)	-0.040** (0.017)	-0.041** (0.018)		-0.039** (0.019)
Education									-0.080*** (0.013)	-0.004 (0.017)
Constant	-0.005 (0.013)	0.001 (0.013)	0.006 (0.013)	-0.006 (0.014)	0.012 (0.014)	-0.089*** (0.015)	-0.071*** (0.017)	-0.069*** (0.018)	-0.015 (0.013)	-0.068*** (0.018)
Observations	5,259	5,259	5,259	4,895	4,895	2,636	2,636	2,454	5,254	2,453
Adj. $R^2$	0.005	0.013	0.014	0.002	0.016	0.004	0.007	0.006	0.007	0.006

Source: BCS. Each column reports the result of a regression analysis with the variable at the top of the column as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figures 1B and 2A for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table 8.16: Regressions of Life Satisfaction on Personality, IQ, Achievement, Grades, and Schooling – BCS

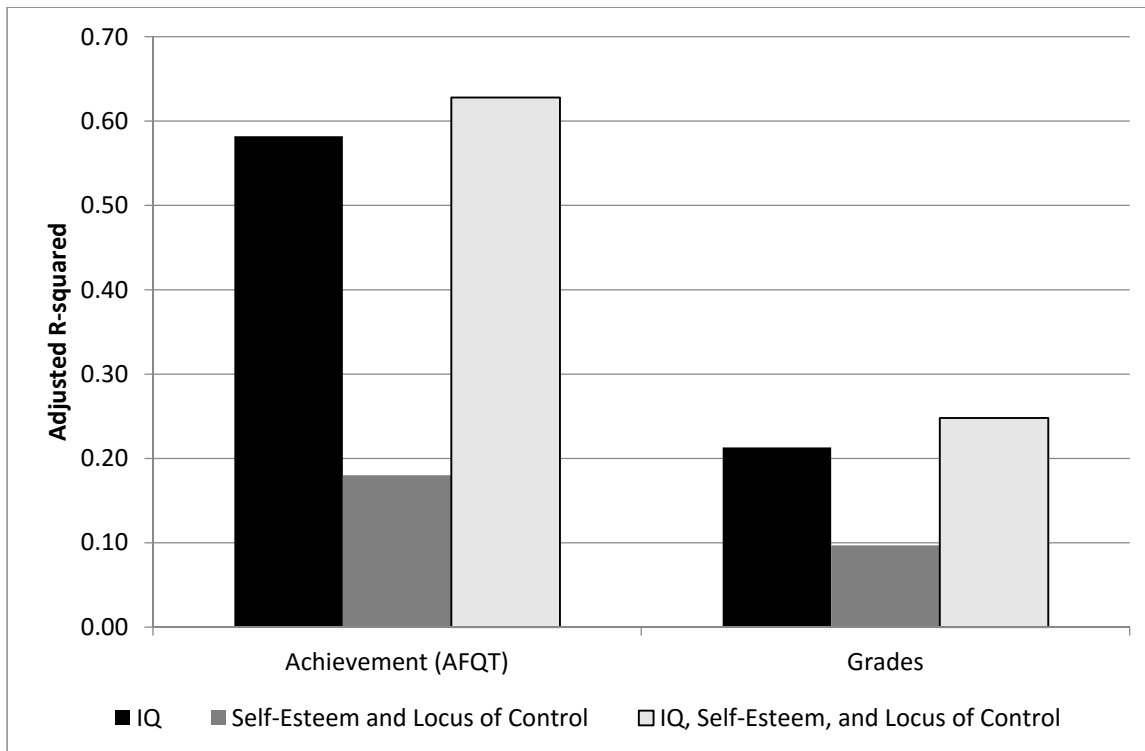
	(1) Life Satisf.	(2) Life Satisf.	(3) Life Satisf.	(4) Life Satisf.	(5) Life Satisf.	(6) Life Satisf.	(7) Life Satisf.	(8) Life Satisf.	(9) Life Satisf.	(10) Life Satisf.
IQ	0.085*** (0.014)		0.030** (0.015)		0.039** (0.017)		0.028 (0.021)	0.050** (0.024)		0.048** (0.024)
Self-esteem		0.058*** (0.015)	0.058*** (0.015)		0.066*** (0.015)		0.073*** (0.020)	0.076*** (0.021)		0.077*** (0.021)
Locus of Control		0.058*** (0.016)	0.051*** (0.017)		0.056*** (0.017)		0.051** (0.023)	0.068*** (0.024)		0.063*** (0.024)
Disorganized		-0.053*** (0.019)	-0.043** (0.020)		-0.037* (0.020)		-0.025 (0.028)	-0.026 (0.029)		-0.023 (0.029)
Anti-social		-0.007 (0.018)	-0.009 (0.018)		-0.018 (0.018)		0.013 (0.026)	0.007 (0.027)		0.006 (0.027)
Neuroticism		0.004 (0.017)	0.005 (0.017)		0.011 (0.017)		0.036 (0.022)	0.043* (0.023)		0.043* (0.023)
Introversion		-0.076*** (0.017)	-0.077*** (0.017)		-0.077*** (0.017)		-0.087*** (0.023)	-0.081*** (0.024)		-0.081*** (0.024)
PCLT				0.061*** (0.014)	-0.013 (0.017)			-0.032 (0.023)		-0.036 (0.023)
Grades						0.110*** (0.018)	0.065*** (0.020)	0.072*** (0.022)		0.060*** (0.023)
Education									0.092*** (0.013)	0.036* (0.020)
Constant	0.014 (0.014)	0.009 (0.013)	0.006 (0.013)	0.019 (0.014)	0.004 (0.014)	0.065*** (0.018)	0.025 (0.020)	0.019 (0.021)	0.027** (0.013)	0.015 (0.021)
Observations	5,258	5,258	5,258	4,894	4,894	2,636	2,636	2,454	5,253	2,453
Adj. $R^2$	0.007	0.028	0.029	0.003	0.030	0.014	0.033	0.038	0.009	0.039

Source: BCS. Each column reports the result of a regression analysis with the variable at the top of the column as the dependent variable and the variables in the rows as the independent variables. All variables are standardized to mean zero and standard deviation 1. See Figures 1B and 2A for a description of the variables. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix 9: Using IQ Measures Obtained at Later Ages of Childhood in NLSY79

The NLSY79 data lacks a single IQ score for all survey members. Instead, it has scores on different IQ tests taken at different ages of childhood. We form an IQ score by equating scores across different exams at common percentiles. Salkever (2015) raises the possible objection that since the ages at which students take the tests varies (from early childhood to the end of high school), the use of early measures of IQ introduces additional measurement error (and lesser predicative power), which would not be present for later life measures. Figure 10.1 shows the Adjusted R-squared when the sample is restricted to those who took IQ tests within 5 years of taking the AFQT test (post-1974). The sample is reduced to 315 observations, but does not qualitatively change the results displayed in Figure 1C in the text. The  $R^2$  attributed to IQ for achievement scores increases (from .48 to .58), but has little effect on the predictive power of grades. The additional adjusted R-squared contributed by personality measures is slightly reduced and the overall adjusted R-squared is somewhat higher, but results are qualitatively similar.

Figure 9.1: NLSY regression of AFQT and Grades on IQ and personality (restricted to those taking IQ tests after 1974)



Notes: see Figure 1C in the text.



## 10: References

- Anderson, Harry E, and Donald A Leton. 1964. "Factor Analyses of the California Test of Mental Maturity." *Educational and Psychological Measurement* 24 (3):513-523.
- Borghans, Lex, Angela L. Duckworth, James J. Heckman, and Bas ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4):972-1059.
- Borghans, Lex, James J. Heckman, Bart H. H. Golsteyn, and Huub Meijers. 2009. "Gender Differences in Risk Aversion and Ambiguity Aversion." *Journal of the European Economic Association* 7 (2-3):649-658.
- Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6):1087-1101.
- Goldberg, Lewis R. 1992. "The Development of Markers for the Big-Five Factor Structure." *Psychological Assessment* 4 (1):26-42.
- Huepe, David, María Roca, Natalia Salas, Andrés Canales-Johnson, Álvaro A. Rivera-Rei, Leandro Zamorano, Aimée Concepción, Facundo Manes, and Agustín Ibañez. 2011. "Fluid Intelligence and Psychosocial Outcome: From Logical Problem Solving to Social Adaptation." *PLoS ONE* 6 (9):e24858.
- Jensen, A. R. 1998. *The G Factor: The Science of Mental Ability*: Westport, Conn.: Praeger.
- Kettner, Norman. 1976. "Armed Services Vocational Aptitude Battery (ASVAB Form 5): Comparison with GATB and DAT Tests: Final Report for Period May 1975-October 1976." Department of Defense, Department of the Air Force, Air Force Systems Command, Air Force Human Resources Laboratory, Brooks Air Force Base, TX.
- Mackintosh, Nicholas J., and E. S. Bennett. 2005. "What Do Raven's Matrices Measure? An Analysis in Terms of Sex Differences." *Intelligence* 33 (6):663-674.
- Maltby, John, Liz Day, and Ann Macaskill. 2010. *Personality, Individual Differences and Intelligence*: Pearson Education.
- Omizo, Michael M. 1980. "The Differential Aptitude Tests as Predictors of Success in a High School for Engineering Program." *Educational and Psychological Measurement* 40 (1):197-203.
- Salkever, David. 2015. "Interpreting the NLSY79 Empirical Data on "IQ" and "Achievement": A Comment on Borghans et al.'s "Identification Problems in Personality Psychology"." *Personality and Individual Differences* 85:66-68.
- Thorndike, R.L., and E.P. Hagen. 1977. *Measurement and Evaluation in Psychology and Education*: John Wiley.
- Wang, Lin. 1993. "The Differential Aptitude Test: A Review and Critique." Presented at the Annual Meeting of the Southwest Educational Research Association, Austin TX.