Maastricht University

What Grades and Achievement Tests Measure

Citation for published version (APA):

Borghans, L., Golsteyn, B., Heckman, J. J., & Humphries, J. E. (2016). What Grades and Achievement Tests Measure. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(47), 13354-13359. https://doi.org/10.1073/pnas.1601135113

Document status and date:
Published: 22/11/2016

DOI:
10.1073/pnas.1601135113

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 25 Apr. 2024

What grades and achievement tests measure

Lex Borghans[a,b], Bart H. H. Golsteyn[a], James J. Heckman[c,d,1], and John Eric Humphries[c]

[a]Department of Economics, Maastricht University, 6200 MD Maastricht, The Netherlands; [b]Research Centre for Education and the Labour Market, Maastricht University, 6200 MD Maastricht, The Netherlands; [c]Department of Economics, The University of Chicago, Chicago, IL 60637; and [d]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.

[1]To whom correspondence should be addressed. Email: jjh@uchicago.edu.

This article contains supporting information online at 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	IQ	Achievement tests	Grades	Personality measures	Adult outcomes
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.[‡] 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 consilience of the evidence across the diverse datasets that justifies the conclusions of this paper.[§]

Table 2. Correlations (Pearson correlations)

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

P values are presented in *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.[¶] 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

[‡]Across datasets, the survey instruments differ somewhat. The definitions are given in *SI Appendix*.

[§]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.

[¶]*SI Appendix* locations of the source regressions for Figs. 1, 2, and 3 are given in the notes of each figure.

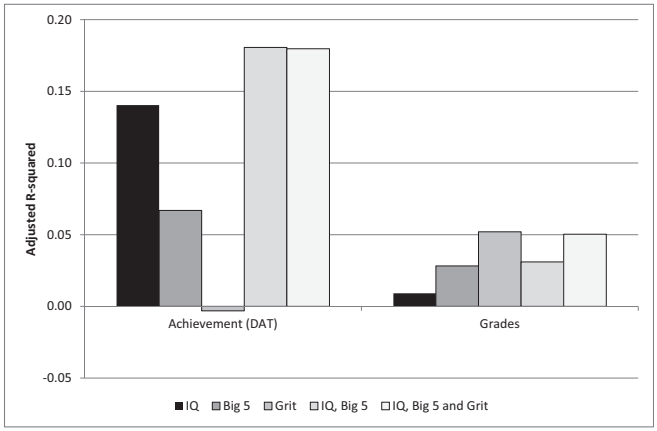

Fig. 1. Decomposing achievement tests and grades into IQ and personality. Stella Maris. The Stella Maris data include 347 Dutch high school students aged 15 or 16 y old in 2008. The figure shows the adjusted R^2 values of two sets of five regressions: DAT/grades on IQ, DAT/grades on the Big Five, DAT/grades 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.[#] However, the variation explained by IQ and personality is relatively small. Consider, for example, the contribution to explained variance from a regression of log wages on IQ, personality, scores on achievement tests, and grades—reported in various combinations. Column 1 in 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

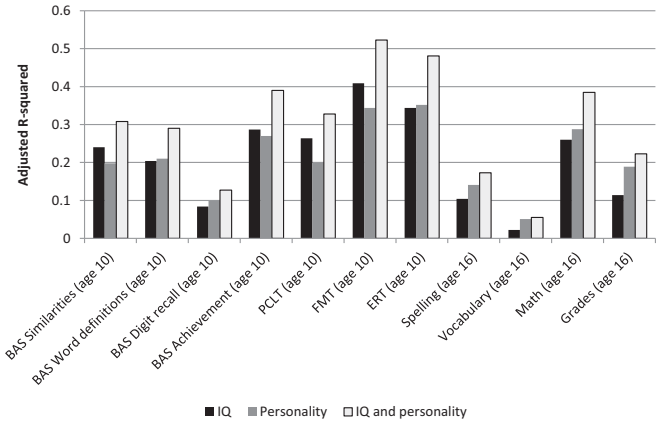

Fig. 2. Decomposing achievement tests and grades into IQ and personality. BCS. The BCS follows a cohort of children born in Britain during one week in April of 1970 until 2016. The sample included 17,198 in 1970. The data contain information collected at age 10 y old on the children's cognitive ability [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.

[#]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.

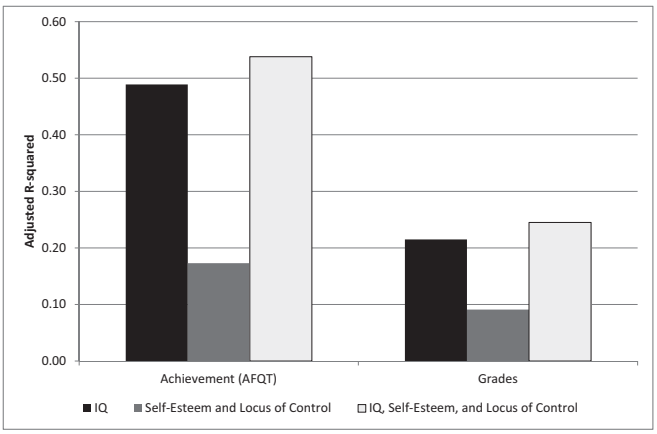

Fig. 3. Decomposing achievement tests and grades into IQ and personality. NLSY79. The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14–22 y old when first surveyed in 1979. The individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. Rotter 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.[||]

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

[||]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).

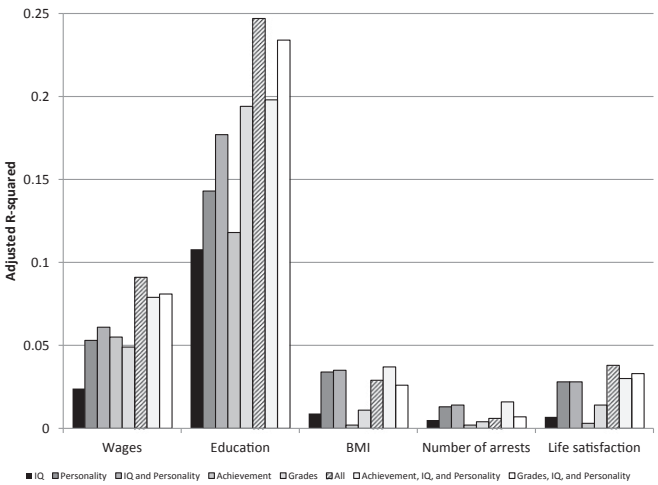

Fig. 4. Decomposing life outcomes into IQ and personality. BCS. Source: BCS 1970 (see Fig. 2). Wages are log wages at age 38 y old. All other measures are measured at age 34 y old and standardized to be mean of zero and SD of one. Education is the nominal age at which a degree is obtained. The figure 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

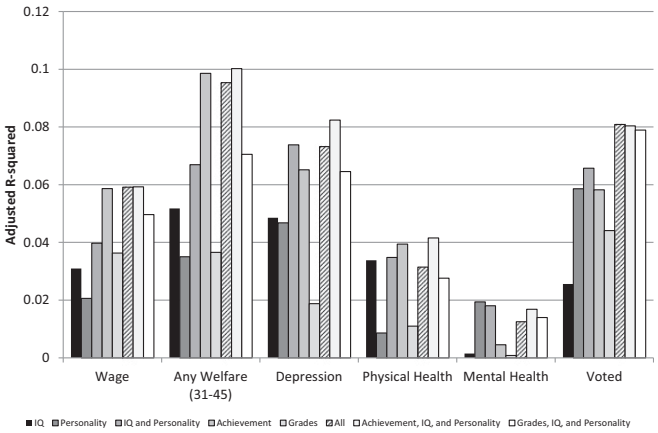

Fig. 5. Decomposing life outcomes into IQ and personality. NLSY79. Outcomes from the NLSY79. All outcomes are at age 40 y old unless otherwise noted. Wages are log wages. Depression is the Center of Epidemiological Studies (CESD) six-item depression scale. Physical health is the SF12 self-reported measure 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.

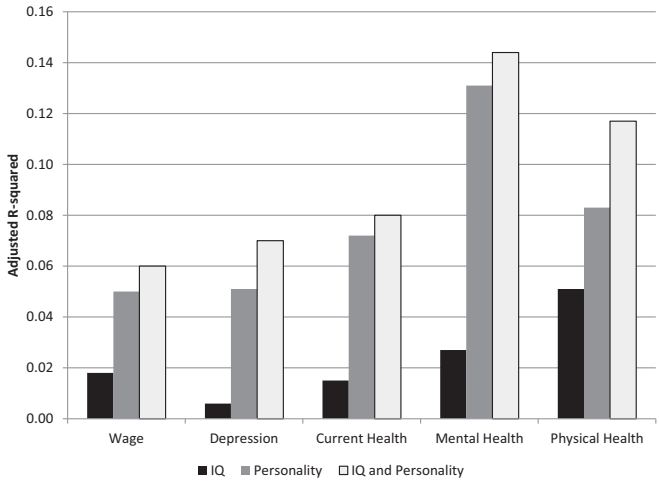

Fig. 6. Decomposing life outcomes into cognition and personality. MIDUS. Data from the MIDUS 1995–1996 and 2004–2006. For privacy, income is reported in 42 unique bins in the MIDUS data. We assign individuals the average of their income bin. Sixty-one individuals in the top bin of $200,000 or higher 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 VIDI 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).

1. Nisbett RE (2009) *Intelligence and How to Get It: Why Schools and Cultures Count* (Norton, New York).
2. Nisbett RE, et al. (2012) Intelligence: New findings and theoretical developments. *Am Psychol* 67(2):130–159.
3. Lindquist EF (1951) Preliminary considerations in objective test construction. *Educational Measurement*, ed Lindquist EF (American Council on Education, Washington, DC), pp 119–158.
4. Heckman JJ, Kautz T (2012) Hard evidence on soft skills. *Labour Econ* 19(4):451–464.
5. 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.
6. 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).
7. 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.
8. Elango S, Hojman A, García JL, Heckman JJ (2016) Early childhood education. *Means-Tested Transfer Programs in the United States II*, ed Moffitt R (Univ of Chicago Press, Chicago).
9. Anastasi A, Urbina S (1997) *Psychological Testing* (Prentice Hall, Upper Saddle River, NJ), 7th Ed.
10. Borghans L, Duckworth AL, Heckman JJ, ter Weel B (2008) The economics and psychology of personality traits. *J Hum Resour* 43(4):972–1059.
11. Borghans L, Meijers H, ter Weel B (2008) The role of noncognitive skills in explaining cognitive test scores. *Econ Inq* 46(1):2–12.
12. Herrnstein RJ, Murray CA (1994) *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York).
13. Murnane RJ, Willett JB, Levy F (1995) The growing importance of cognitive skills in wage determination. *Rev Econ Stat* 77(2):251–266.
14. Hanushek EA, Woessmann L (2008) The role of cognitive skills in economic development. *J Econ Lit* 46(3):607–668.
15. 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.
16. 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.
17. Segal C (2012) Working when no one is watching: Motivation, test scores, and economic success. *Manage Sci* 58(8):1438–1457.
18. 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.
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.