

Conscientiousness, Extraversion, College Education, and Longevity of High-Ability Individuals*

Peter A. Savelyev

The College of William & Mary and IZA

June 10, 2020

*A version of this paper published by the *Journal of Human Resources* is available from <https://doi.org/10.3368/jhr.58.1.0918-9720R2>. This paper was presented to The Health Economics Workshop at the NBER Summer Institute; The Annual Meeting of the American Economic Association in Chicago; The European Economic Association and Econometric Society Annual European Meeting in Gothenburg, Sweden; The Health Economics Workshop and the Labor Group Seminar at the University of Chicago; Applied Microeconomics seminar at the University of North Carolina; The Institute on Health Economics, Health Behaviors, and Disparities at Cornell University; Economics seminar of the Andrew Young School of Policy Studies, Georgia State University; Empirical Micro seminar of the University of Houston and Rice University; RAND Labor and Population Seminar in Santa-Monica; Departmental Seminar in Economics at Vanderbilt University; the Applied Microeconomics Seminar at Vanderbilt Law School; The Quantitative Methods Brown Bag at the Department of Psychology and Human Development, Peabody College, Vanderbilt University; David Eccles School of Business research seminar, University of Utah; and the Upjohn Institute for Employment Research seminar. I thank participants of these meetings for useful suggestions and stimulating discussions. I am grateful to Gary Becker, Gabriella Conti, Miriam Gensowski, Mike Grossman, Tim Kautz, Don Kenkel, Adriana Lleras-Muney, Willard Manning, David Meltzer, Rémi Piatek, Ben Williams, and especially Jim Heckman for productive comments. This version of the paper directly benefited from research assistance provided by Atticus Bolyard and Max Sacher and from proofreading by Xiaoyu (Nancy) Chen, Renee Garrow, Isabel Harber, and Eli Rothleder. Work on this paper stretched over a decade, over which many RAs contributed to its development, namely Mihir Gandhi, Kai Hong, Hanchen Jiang, Adam Shriver, Ivana Stosic, and Keron Tan. Summer interns Keith Dent, Son Nghiem, John Spraul, and Cody Vaughn used previous versions of this paper to learn empirical skills and provided help. The Terman data are provided by the Interuniversity Consortium for Political and Social Research, Ann Arbor, MI. Peter Savelyev gratefully acknowledges research support from the College of William & Mary, Vanderbilt, and the ERC at the University of Chicago. An early version of this research was supported by the Merck Quantitative Science Graduate Fellowship in Health Economics. The views expressed in this paper are those of the author and may not coincide with those of the funders.

Abstract

Using the 1922–1991 Terman Life-Cycle Study of Children with High Ability, I investigate the relationship between childhood noncognitive skills, college education, and longevity of a high-IQ population and find a strong relationship between college education and longevity for men. Conscientiousness and Extraversion are strongly related to longevity of men, even though their effects on education are, at best, weak. I demonstrate a number of behavioral mechanisms behind the estimated effects on longevity. I also find that men with higher levels of education and skills have superior health over the lifespan. For women of this historical cohort (born around 1910), who live at least as long as college-educated men, I find no evidence of a relationship between education, noncognitive skills, and longevity.

Key words: noncognitive skills, college education, longevity, human potential-education matching, compression of morbidity, sex difference, Big Five personality taxonomy, Terman Life-Cycle Study of Children with High Ability

JEL codes: I12, J24

1 Introduction

It is controversial in the literature whether education affects longevity. It is also unclear whether this relationship, if it exists, fades with greater cognitive skills. In this paper I find that there is a strong relationship between college education and longevity for high-ability men, which contradicts a claim in the literature that education is only productive for health at low levels of IQ. Moreover, I suggest a new mechanism, which I call “human potential-education matching,” through which IQ can increase the effect of education on health. I also find strong relationships between the well-established noncognitive skills of Conscientiousness and Extraversion and male longevity. This finding confirms the known link between Conscientiousness and longevity for high-ability individuals after accounting for a number of potential sources of bias. However, the strong link between Extraversion and longevity in a high-ability population is a new finding. I support these conclusions by demonstrating likely behavioral and socioeconomic mechanisms behind the estimated relationships. I also find that the confounding role of noncognitive skills on the education-longevity gradient is negligible for a high-ability population.

I use the Terman Life-Cycle Study of Children with High Ability ([Terman, 1986](#)), which follows about 1,500 high-IQ men and women from California born around 1910. The dataset is well-suited for the study of developmental origins of longevity since it contains a unique combination of measures: childhood IQ and noncognitive skills, detailed family background characteristics, and 70 years of prospective observations that include education, health behaviors, health, and mortality. Datasets like Terman’s that combine the availability of quality early skill measures with a follow-up long enough to study longevity and its mediators are rare.

I jointly estimate a system of equations that includes an ordered logit model of schooling, a mixed proportional hazard model of mortality, and equations linking a low-dimensional set of latent noncognitive skills to their noisy measures. On top of controlling for a detailed set of background variables and IQ, I also control for a set of latent noncognitive skills, which are linked to the established Big Five taxonomy. In addition, I control for unobserved heterogeneity modeled via a semi-parametric heterogeneity model ([Heckman and](#)

Singer, 1984).¹ Controlling for unobserved heterogeneity affects both qualitative and quantitative results. The model is supported by placebo tests using eight early health measures as outcomes.

This paper makes several contributions. First, in the health economics literature, even though education and longevity strongly correlate, the causal status of this relationship is still controversial and little is known about the effect of college education on longevity (see Section 2 for literature review). This paper supports claims about the causal relationship between education and longevity for men by estimating both the relationship and plausible mechanisms behind it. Women of this cohort live at least as long as college-educated men, but I find no effect of education on their longevity.

Second, I contribute to understanding whether the relationship between education, health, and longevity still holds for individuals with high cognitive ability. If cognition is a perfect substitute for education in health production, then education might only be productive for the health of people with low cognitive ability, as long as education has diminishing health returns. Using parental education as an instrumental variable for education, Auld and Sidhu (2005) find for a pooled sample of men and women that years of education are beneficial for health, but only for low-IQ individuals and at low levels of schooling. Conti and Heckman (2010) study the effects of education on health-related outcomes of the general population and find that cognitive ability tends to increase beneficial effects of education for men but decrease such effects for women. This paper finds no effect of college education on longevity for high-IQ women but sizable effects for men. Relative to high school education, some college education increases life expectancy of men at age 30 by five years, and a 4-year college degree or above adds ten years. Therefore, my result for women is consistent with both Conti and Heckman (2010) and Auld and Sidhu (2005), but the result for men is consistent only with Conti and Heckman (2010) and contradicts Auld and Sidhu (2005).

Moreover, I provide evidence for a mechanism, which I call *human potential-education matching*, through which higher IQ can potentially increase the effect of education, in con-

¹The approach is also known as a “latent class model” (Aitken and Rubin, 1985), a “discrete factor approximation model” (Mroz, 1999), and a “discrete factor random-effects model” (Gilleskie, 2014).

trast to the possible substitution effect discussed above. In the case of this population, high-IQ individuals with low levels of education could be dissatisfied with their jobs and lives and thus invest less in their health or be involved in substance abuse or both. Substance abuse can be viewed as self-treatment of depression and other psychological issues (Darden and Papageorge, 2018). This mechanism can be viewed as a special case of value of life mechanism suggested by Cutler and Lleras-Muney (2008). I provide evidence for Terman men consistent with this hypothesis. First, white high school graduates with high ability have notably larger mortality than white males from the general population of the same cohort despite having both superior average ability and education. A greater mismatch between ability and education for low-educated Terman subjects compared to an average man from the general population would explain this paradox. Second, Terman subjects with low levels of education are more likely to report a dislike of their current occupation, lack of control in the choice of work, lack of intellectual challenge, and a hindered life due to inadequate education. They are also more likely to report higher likelihood of heavy drinking and greater difficulties with mental and general health. Among other mechanisms known from the literature, this suggested mechanism could be behind the estimated sizable effects of education on longevity.²

Third, a growing literature in the economics of human development shows the importance of noncognitive skills for generating essential life outcomes including health behaviors and health (e.g., Conti et al., 2010; Heckman et al., 2018). Noncognitive skills are also referred to as “socioemotional skills,” “soft skills,” or “personality traits,” and are defined as patterns of thought, feelings, and behavior (Borghans et al., 2008). I call these characteristics “skills” rather than “traits” following the economic literature that highlights their malleability, especially in childhood and young adulthood (Heckman and Kautz, 2014). One reason for their high predictiveness is that they can be linked to essential parameters of the utility function, namely time preference and risk aversion (see Almlund et al., 2011 for a survey). Another reason is that they capture important information about skills, such as perseverance and sociability, that are either not captured by traditional economic variables or captured only indirectly and partially.

²I discuss possible mechanisms of education effect on longevity in detail in Section 5, including a more detailed discussion of the suggested matching mechanism.

Noncognitive skills can affect health outcomes directly. For instance, conscientious patients are better at following complex diabetes treatment instructions (e.g., [Vollrath et al., 2007](#)). They can also affect health outcomes indirectly through education, other skills, and the effectiveness of health production. Skills can also interact with causes of health behaviors, such as peer pressure ([Hsieh and van Kippersluis, 2018](#)).

While there are multiple taxonomies to describe noncognitive skills, I use the Big Five Taxonomy of Personality, one of the most established classification methods ([John and Srivastava, 1999](#)). This comprehensive taxonomy includes the personality traits of Conscientiousness, a propensity to be thoughtful about the future and to follow rules; Openness, a propensity to be curious about new ideas and experiences; Extraversion, a propensity to be active towards social and material worlds; Agreeableness, a propensity to be friendly to others; and Neuroticism, a measure of emotional instability. The Big Five taxonomy had not yet been established at the time when Terman data were collected, but many essential measures of personality related to the Big Five had been established. A strong link between Terman personality measures and the Big Five traits has been demonstrated by psychologists ([Martin and Friedman, 2000](#)).³

This paper contributes to the literature in the economics of human development, which acknowledges that noncognitive skills are multi-dimensional (e.g., [Borghans et al., 2008](#)). However, many studies rely on a one-dimensional noncognitive factor often due to a lack of sufficiently detailed early-life measures of noncognitive skills in available datasets (e.g., [Cobb-Clark et al., 2014](#); [Conti et al., 2010](#); [Heckman et al., 2018, 2006](#)). This paper accounts for multi-dimensional and correlated noncognitive factors and finds that their effects on longevity and health-related outcomes differ.

Fourth, an established strand in the literature focuses on the role of skills as confounding factors of the effect of education on health-related outcomes and on the size of the ability bias. The size of the ability bias found in the literature varies paper to paper, as does the definition of skills. For instance, [Van Der Pol \(2011\)](#) studies the ability bias due to omitted time preference and risk attitude controls and finds a modest 8% upward bias conditional on other controls. In contrast, [Heckman et al. \(2018\)](#) use cognitive and GPA measures

³See Web Appendix A for more information about the Big Five taxonomy and Terman's noncognitive measures.

as proxies for ability and find a large upward ability bias of 20–80% for health-related outcomes conditional on background controls.⁴ I supplement results from this literature by finding no ability bias for the population of high-ability individuals despite having access to a comprehensive set of noncognitive skills. This result is explained by the lack of a strong relationship between noncognitive skills and education in the population with extraordinarily high IQs.

Fifth, much research has been done on the association between personality and longevity (see [Chapman et al. \(2011\)](#) for a survey). Among the Big Five traits, only Conscientiousness is consistently associated with longevity across studies. Previous research by psychologists based on the Terman data finds a positive association between longevity and Conscientiousness for men and for a pooled sample of men and women ([Friedman et al., 1993](#)). The authors condition their Cox proportional hazard regressions on one or several additional psychological traits. I show that this result is preserved when I control for a large set of essential background variables, IQ, and unobserved heterogeneity.

I also contribute to understanding the role of childhood Extraversion by finding a strong effect on longevity that is comparable to the effect of childhood Conscientiousness.⁵ Previous studies report no effect of Extraversion based on the Terman data ([Friedman, 2000](#)), possibly due to controlling for neither measurement error in Extraversion measures nor for observed or unobserved heterogeneity.⁶ Openness shows no precisely-determined effects on longevity. Finally, as in the previous literature based on the same data, I do not find any statistically significant effects of Agreeableness and Neuroticism. I supplement the estimated effects on longevity with the analysis of possible behavioral mechanisms behind these effects.

Finally, I find positive effects of education and skills on health over the lifespan, which are consistent with the highly policy-relevant but controversial *compression of morbidity* hypothesis. Compression of morbidity implies an increasing share of the lifespan spent in

⁴I refer to approximate percentage differences between light grey and dark gray bars in Figure 2 of [Heckman et al. \(2018\)](#) for panels describing daily smoking and health limitations to work.

⁵Terman’s available measures of early Extraversion largely represent its sociability facet.

⁶Previous studies labeled this factor as “Sociability.”

good health as longevity increases, according to an approach by [Cutler et al. \(2014\)](#).⁷

2 Relating this Paper to Literatures in Health Economics and Epidemiology: Methodology and Results

This section complements the introduction by elaborating on the relationship of this paper to literatures in health economics and epidemiology. I cluster literature results by methodological approach, as each approach comes with specific interpretations and limitations. See also [Grossman \(2015\)](#) and [Galama et al. \(2018\)](#) for detailed recent literature reviews on the effect of education on health-related outcomes.

Compulsory Schooling Laws A popular method of establishing the effect of education on health and longevity is using changes in compulsory schooling laws as an instrument for years of schooling. Results based on this approach differ: [Lleras-Muney \(2005\)](#) finds a large effect of schooling on mortality for the US, but [Mazumder \(2008\)](#) shows that these effects are not robust to including state-specific time-trends. Similarly, analyses by [Meghir et al. \(2018\)](#) and [Albouy and Lequien \(2009\)](#) do not find a beneficial effect of schooling on longevity for Sweden and France, respectively. In contrast, [van Kippersluis et al. \(2011\)](#) use Dutch compulsory schooling laws and find a beneficial effect on mortality for men.⁸

These disagreements may be related to possibly different effects of education for different countries and cohorts. Issues with statistical power and weakness of instruments are other possible explanations. In fact, at least for France and the US, the instrument appears weak ([Galama et al., 2018](#)). Other limitations of the method include possible confounding effects from other progressive reforms, such as public health improvements and school lunch programs, and relying on identification of the effect for a specific group of marginal dropouts from a compulsory school. The population of marginal school dropouts may have a wide range of IQs but can be expected to have lower IQ levels on average. My paper

⁷Depending on available data and the research question, “good health” in this definition could mean, for instance, “not disabled” or “free from chronic disease.”

⁸Some papers investigated not longevity, but health-related outcomes. For instance, [Clark and Royer \(2013\)](#) and [Arendt \(2005\)](#) do not find a causal effect of schooling on health for the UK and Denmark respectively.

complements these results by focusing on a different education level (college education) identified for a different population (high-ability people).

Draft Avoidance Unlike the effect of compulsory education, the effect of college education on longevity is largely unexplored, perhaps since suitable natural experiments are less readily available. An exception is a paper by [Buckles et al. \(2016\)](#), which uses the state-by-cohort-level mortality rates of men and identifies a strong and statistically significant effect of college graduation on longevity from Vietnam War draft avoidance. Limitations of the [Buckles et al.](#) paper include that it identifies the effect only for those who choose more education as a result of the risk of draft and that it is difficult to separate the effect of college from the effect of being drafted to serve in Vietnam.

My paper further explores the effect of college education on longevity, and complements the paper by [Buckles et al.](#) in a number of ways. First, I use a different methodology discussed below in this section. Second, I study a different population and my sample includes women. Third, rather than looking at binary mortality over a 25-year period, I study mortality in continuous time over 56 years of life, which allows me to estimate the survival function and, as a result, the effect on life expectancy and the value of remaining life.

Twin Studies This literature uses twin fixed effects as a source of identification. Studies typically use the total number of years of schooling to maximize statistical power, and so they measure the average effect of schooling across various education levels. Very few twin studies are long enough and have a large-enough sample size to reliably estimate the effect of education on longevity. Results based on these few studies are mixed.

In particular, based on Swedish twin data [Lundborg et al. \(2016\)](#) find strong effects of education on longevity for both men and women. Results based on Danish twin data vary: [Madsen et al. \(2010\)](#) and [Behrman et al. \(2011\)](#) find no effects, while [Van Den Berg et al. \(2015\)](#) finds an effect on mortality for men but not for women based on a different methodology. [Savelyev et al. \(2020\)](#) use US Minnesota Twin Registry (MTR) data and demonstrate a statistically significant effect of schooling on mortality for men.⁹ Again,

⁹[Amin et al. \(2015\)](#) also use MTR data, which they combine with the Mid-Atlantic Twin

these different results may be partly related to differing effects of education for different countries.

Limitations of twin studies include a possibility of unobserved confounders behind estimated effects; large attenuation bias due to measurement error in education, which is amplified by taking differences across twins; and difficulties in generalizing results outside of twin populations.

This paper complements results of this literature by adding additional evidence for longevity in the US, which we know very little about from twin studies (Savelyev et al., 2020). Moreover, I study a specific policy-relevant education margin (college education) rather than one year of schooling averaged across different margins.

Structural Equation Modeling An alternative approach to causal inference is to explicitly model the interdependence between observed and unobserved confounders, education, and health-related outcomes (e.g., Conti and Heckman, 2010). I complement these results by studying longevity, an objective outcome that incorporates all observed and unobserved inputs to the health production function and a unique margin for the quantity rather than quality of life (Becker, 2007; Galama and van Kippersluis, 2015; Murphy and Topel, 2006). As I find no selection into education level on noncognitive skills in this paper whereas Conti and Heckman (2010) find strong selection, we can conclude that selection on noncognitive skills strongly depends on the range of IQ. I also extend the methodology by accounting for multidimensional latent noncognitive skills combined with modeling unobserved heterogeneity that is not accounted for by latent noncognitive skills and find that ignoring such unobserved heterogeneity leads to a sizable bias.¹⁰

Following the structural equation approach, Bijwaard, van Kippersluis, and Veenman (2015) (BKV) and Bijwaard et al. (2015) study the relationship between IQ, education, and mortality. The authors use data from the Netherlands and find an effect of education on

Registry (MATR). The authors concentrate on health-related outcomes, not longevity, and find little evidence of the causal effect of schooling. However, Savelyev et al. (2020) argue that pooling sexes and using noisy outcomes masks important relationships.

¹⁰The longevity results of this paper are in line with a number of papers that identify effects of post-compulsory education on health-related outcomes other than longevity (Bolyard and Savelyev, 2020; Conti et al., 2010; Currie and Moretti, 2003; de Walque, 2007; Grimard and Parent, 2007; Heckman et al., 2018).

mortality and a strong selection into education based on IQ. BKV extend the framework by [Conti and Heckman \(2010\)](#) by accounting for left truncation. The authors note that they could not incorporate a model for noncognitive skills into their analysis due to data limitations, which could possibly bias their results. In contrast, my model incorporates multidimensional noncognitive skills. After showing that left-truncation issues are negligible for the Terman data, I extend the framework outlined by [Conti and Heckman \(2010\)](#) by accounting for unobserved heterogeneity that is not captured by latent noncognitive skills and show that the lack of this control leads to a large bias. This paper complements that of BKV not only methodologically, but also substantively. First, I study a different population from a different country.¹¹ Second, while the authors look at the effect of completing any schooling above primary school, I examine the effect of college education. Third, while BKV emphasize the effect of IQ, I concentrate on the effects of noncognitive skills conditional on high IQ.

A related paper by [Hong, Savelyev, and Tan \(2020\)](#) based on the Wisconsin Longitudinal Survey data also suggests an effect of college education on longevity for men, but that paper is based on a different population (high school graduates from Wisconsin) and focuses on estimating contributions of multiple mechanisms of longevity to the total effect on mortality, something that I could not do in this paper due to sample size and data limitations. Unlike this paper, the [Hong, Savelyev, and Tan \(2020\)](#) paper studies neither effects of the Big Five nor their confounding role.

Limitations of structural equation methods include reliance on additional structural assumptions and generalizations of the conditional independence assumption. However, using semi-parametric models and accounting for unobserved heterogeneity, as I do in this paper, makes these assumptions more plausible. Moreover, my assumptions are consistent with placebo tests.

Propensity Score Method [Bijwaard and Jones \(2019\)](#) use the propensity score method and find that education hardly affects mortality for men born in 1944–1947 in the Netherlands. In a separate paper, [Bijwaard et al. \(2017\)](#) estimate a case-specific education gradient in mortality by combining the propensity score method with family fixed effects. The authors

¹¹The authors study a cohort from the province of Noord-Brabant in the Netherlands.

use data on men with brothers from the Swedish Military Conscription agency (1951–1983) and find that improving education by one level leads to three to 10 additional months of life between ages 18 and 63 and that most of the mortality gains are attributable to mortality from causes that are unrelated to disease (traffic accidents, suicides, etc.). The main limitation of these papers is their reliance on the conditional independence assumption.¹² Substantively, my paper complements these papers by estimating the effects of both education and noncognitive skills and studying a very different population.

Epidemiological Studies of Schooling, IQ, and Longevity Batty and co-authors use Swedish conscripts data and show that IQ is associated with all-cause mortality and with mortality from heart disease and suicide, and that controlling for education markedly attenuates these results (Batty et al., 2009). Also, Batty et al. (2007) show that adolescent IQ is robustly related only to skin cancer. Finally, Batty et al. (2009) show that IQ has a preventative effect on injury mortality. These results are in line with the rest of the literature, which shows that IQ is negatively associated with mortality and that controlling for education attenuates the association (see Batty et al. (2007) and Calvin et al. (2011) for surveys.)

Another source of attenuation is family and genetic influence. Næss et al. (2012) study the confounding effect of a family factor shared by siblings on the association between education and case-specific mortality in adulthood and confirm that shared family factors confound the effect, with varying magnitudes. The authors used data on all Norwegians born in 1940–1959 who had one or more siblings within the cohort and were alive in 1990.

Given that, based on epidemiological studies, IQ is expected to be a major confounder, it is useful that I am able to control for it twice: (1) by using a high-IQ sample and (2) by using early IQ as a control. My paper complements these studies by estimating the effect of education and socioemotional skills conditional on high IQ.

3 Terman Data

The Terman Study started in 1922 and continued through 1991. The sample consists of about 1,500 men and women selected in their childhood and adolescence for their high

¹²The authors call it “unconfoundedness assumption.”

ability based on teachers' nomination followed by an IQ test with a cut-off value of 140.¹³ The subjects, who are white and mostly financially comfortable, were born, on average, in 1910.¹⁴ The study has an attrition rate below 10%, which is exceptionally low for a 70-year-long prospective study. Moreover, the attrited subjects are known not to differ systematically in terms of education, income, and demographic factors from the main sample (Sears, 1984). There is also no evidence that members of the attrited group differ significantly from others on measures of personality (Friedman et al., 1993).

Table 1 shows descriptive statistics for IQ, education, and 23 background variables. The background variables are grouped into five categories: early health, early childhood investments, parental background, World War II experience, and cohort dummies.

Participants have an average IQ of about 149. As for education, 69–73% of participants earned a bachelor's degree, and 16–20% have some college education below a bachelor's degree. It may seem surprising by modern standards that as many as 10–11% of high-ability subjects did not continue their education beyond high school. The education choice becomes more understandable once we look at historic trends and see that a high school degree for this generation corresponds rank-wise to a college degree for recent cohorts, while only 5% of this population completed college.¹⁵

One important benefit of the longitudinal nature of the Terman study is the detailed education data collected several times from 1922 to 1968, both prospectively and retrospectively. This paper uses all available information to infer the highest education level, an approach that contributes to the accuracy of education measurement.

Table 2 shows health-related outcomes except for longevity, which is described later in

¹³Teachers nominated from one to five children, usually four, from classes of 30–50 pupils. Teachers were asked to base nominations on such proxies of high IQ as intelligence, quickness of grasp, originality, ability to reason clearly about new and difficult problems, breadth and accuracy of information, command of language, common sense, and independence of judgment. Conditional on talents, younger age was viewed favourably (Terman et al., 1925).

¹⁴Terman et al. (1925) refer to the economic status of a majority of families as “fairly comfortable,” and indicates that only a few families were “truly in poverty.” I control for *adequate finances* and *social position* in all models.

¹⁵According to Figure 5 of Heckman and LaFontaine (2010), the high school completion rate for the 1910 general population cohort is about the same as college graduation rate for the 1980 cohort: 30%. Only about 5% of the 1910 cohort finished college.

Table 1: IQ, Education, and Background Variables

Variable	Year of measurement	Males		Females	
		Mean	Standard deviation	Mean	Standard deviation
General Intelligence					
IQ ^(a)	1922	149.2	(10.50)	148.5	(10.24)
Highest Education Level					
High school graduate	1922–1968	0.102	(0.303)	0.112	(0.315)
Some college	1922–1968	0.164	(0.370)	0.200	(0.400)
Bachelor's degree or above	1922–1968	0.734	(0.442)	0.688	(0.463)
Early Health					
Normal birth ^(b)	1922	0.603	(0.489)	0.652	(0.476)
No breastfeeding ^(b)	1922	0.106	(0.308)	0.089	(0.285)
Health rating in 1922 ^(c)	1922	8.545	(1.972)	9.023	(1.906)
Low birthweight (below 2.5 kg) ^(b)	1922	0.021	(0.145)	0.046	(0.207)
Mother's poor health during pregnancy ^(b)	1922	0.168	(0.374)	0.177	(0.382)
Nutrition poor or fair in 1922	1922	0.086	(0.281)	0.072	(0.259)
Severe acute and chronic illnesses since 1922	1928	0.120	(0.324)	0.092	(0.290)
Signs of nervousness	1928	0.206	(0.405)	0.224	(0.416)
Early Educational Investments					
Logarithm of the amount of parental tutoring ^(d)	1922	0.445	(0.370)	0.410	(0.358)
Logarithm of the amount of private tutoring ^(d)	1922–1928	0.103	(0.373)	0.345	(0.597)
Parental Background					
Father is deceased by 1922	1922	0.085	(0.279)	0.076	(0.265)
Parents are divorced by 1922	1922	0.049	(0.217)	0.050	(0.217)
Father has at least a bachelor's degree	1922	0.330	(0.470)	0.287	(0.453)
Mother is employed	1922	0.136	(0.344)	0.141	(0.348)
Father is a professional	1922	0.254	(0.435)	0.292	(0.455)
Either parent from Europe	1922	0.230	(0.421)	0.213	(0.410)
Parental finances adequate	1922	0.370	(0.483)	0.380	(0.486)
Parental social position below average	1922	0.252	(0.434)	0.153	(0.361)
World War II Experience					
WWII participation	1945	0.435	(0.496)	0.026	(0.158)
WWII combat experience	1945	0.150	(0.356)	0.004	(0.063)
Cohort					
Cohort: 1904–1907	1922	0.243	(0.429)	0.174	(0.378)
Cohort: 1908–1911	1922	0.463	(0.499)	0.468	(0.499)
Cohort: 1912–1915	1922	0.294	(0.455)	0.358	(0.480)
Estimation Sample Size		695		536	

Notes: ^(a)The best estimate of IQ in 1922 is provided by survey organizers and is based on all available test scores including Stanford Binet and Terman Group Tests. ^(b)Indicators of conditions at birth are reported retrospectively by parents in 1922. ^(c)An average over non-missing values of teachers' and parents' ratings is used. Both teachers and parents rated general health of students on a continuous scale by drawing a mark anywhere on a line segment (see Figure A-1 of the Web Appendix). Survey organizers quantified the marks using a scale from 1 to 13. ^(d)Amount of parental and private tutoring is calculated for ages 2–7.

Table 2: Health-Related and Life Satisfaction Outcomes, Males

Variable	Year of measurement	Full sample		High school		Some college		College or above		# of observations
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	
Health-related outcomes and health										
Ever involved in heavy drinking	1940–60	0.415	(0.493)	0.485	(0.503)	0.477	(0.502)	0.392	(0.489)	679
Married once and still married	1960	0.575	(0.495)	0.515	(0.503)	0.455	(0.500)	0.611	(0.488)	684
# of memberships in organizations	1960	3.129	(2.095)	2.262	(1.740)	1.929	(1.731)	3.432	(2.094)	498
Professional occupation ^(a)	1960	0.469	(0.499)	0.104	(0.308)	0.162	(0.370)	0.585	(0.493)	684
Mental difficulty in 1950	1950	0.281	(0.450)	0.321	(0.471)	0.309	(0.465)	0.271	(0.445)	620
Mental difficulty in 1960	1960	0.295	(0.457)	0.318	(0.471)	0.267	(0.445)	0.300	(0.459)	532
General health rating in 1950	1950	4.439	(0.691)	4.375	(0.822)	4.289	(0.889)	4.477	(0.621)	624
General health rating in 1972	1972	4.326	(0.853)	3.741	(1.163)	4.194	(1.084)	4.400	(0.751)	423
General health rating in 1986	1986	4.046	(0.827)	3.533	(0.990)	4.056	(1.040)	4.019	(0.895)	304
Life Satisfaction outcomes										
Definitely chosen life's work ^(b)	1940	0.690	(0.463)	0.565	(0.501)	0.640	(0.483)	0.713	(0.453)	568
Chose, not drifted into occupation ^(c)	1940	0.691	(0.462)	0.468	(0.504)	0.500	(0.503)	0.756	(0.430)	525
Work provides great satisfaction ^(d)	1950	0.763	(0.426)	0.625	(0.490)	0.743	(0.440)	0.780	(0.415)	532
Inadequate education hindered life ^(e)	1950	0.155	(0.362)	0.700	(0.466)	0.360	(0.485)	0.063	(0.242)	368
Has outlets for mental capabilities ^(f)	1950	0.838	(0.369)	0.684	(0.471)	0.794	(0.408)	0.860	(0.348)	500
Lived up to intellectual abilities ^(g)	1950	0.914	(0.280)	0.805	(0.401)	0.930	(0.258)	0.923	(0.268)	538
Enjoys present vocation ^(h)	1950	4.290	(0.883)	4.049	(1.071)	4.296	(0.932)	4.312	(0.853)	535
Early occupation choice fulfilled ⁽ⁱ⁾	1950	0.648	(0.478)	0.310	(0.471)	0.518	(0.504)	0.694	(0.461)	458

Notes: ^(a)Coded based on reported occupation: university faculty, engineer, lawyer, physician, journalist, and the like. ^(b)Definitely chosen life work? 0, no or indefinite, 1, yes. ^(c)Occupation chosen or drifted into. 0, drifted into, 1, definitely chosen. ^(d)Work is among aspects of life that provide the greatest satisfaction. 0, not checked, 1, single or double check. ^(e)Factor hindering to life accomplishments was inadequate education. 0, not checked, 1, single or double check. ^(f)Does life offer satisfactory outlets for mental capabilities? 0, no, 1, fairly well or yes. ^(g)Lived up to intellectual abilities? 1, a total failure, 2, consider my life largely a failure, 3, far short of it, 4, considerably short of it, 5, reasonably well, 6, fully. ^(h)Feelings about present vocation. 1, strongly dislike and hope to change, 2, discontented but will probably stick it out, 3, no serious discontent, 4, fairly content, 5, deep satisfaction and interest. ⁽ⁱ⁾Was early occupation choice fulfilled? 0, no, 1, yes.

the results section,¹⁶ and life satisfaction outcomes by education level.

I use these variables to understand the mechanisms of longevity production and the persistence of the effects on health over the lifespan. Some likely mechanisms, such as smoking or physical exercise in early and mid-life, are missing in the data. However, information is available on self-reported heavy drinking, marriage, memberships in organizations (a proxy for social connections and an active lifestyle), belonging to a professional occupation, and health measures: mental issues and general health ratings. Table 2 demonstrates the same relationships as we will see above from regression analysis: better education corresponds to healthier lifestyles, superior health, and better life satisfaction.

I restrict the data based on a number of criteria chosen prior to estimation. I exclude subjects who: (1) were not born in the period 1904–1915;¹⁷ (2) have missing 1922 ratings of noncognitive skills; (3) have missing education information; (4) dropped out of high school;¹⁸ and (5) died or attrited before age 30.¹⁹ Criteria (1) and (2) have also been used by psychologists investigating the same data (Martin et al., 2007). The resulting estimation sample size is 695 for men and 536 for women.

Measuring Noncognitive Skills Although there are various ways to define noncognitive skills, one of the most established methods is to use the Big Five taxonomy of personality (John and Srivastava, 1999). The data on personality collected in 1922 and 1940 by Terman and coworkers are both theoretically and empirically close to the Big Five taxonomy (Martin et al., 2007).

Table 3 summarizes measures that I use for the main model of this paper.²⁰ The table

¹⁶See Kaplan-Meier survival curves in Figure 1.

¹⁷This restriction makes the cohorts more comparable by excluding a small number of respondents in the tails of the year of birth distribution.

¹⁸Conditional on all other sample restrictions listed here, only two men and five women dropped out of high school. The number of dropouts is so small that reliable estimation of the dropout effect is impossible. Alternatively, it would be a strong assumption to view dropouts and high school graduates as belonging to the same education category.

¹⁹Deaths before age 30 are rare for Terman survey participants, but counting them may bias our estimates towards the reverse causal case. However, as I show above, results are robust to the choice of the truncation point.

²⁰As in previous research based on the same data (Friedman et al., 2010, 1995, 1993), I average teacher's and parent's ratings of children's noncognitive skills to account for all available sources of information. This approach is in line with Murray et al. (2007), who conclude that obtaining ratings from multiple informants is critical for obtaining a full

Table 3: Childhood Noncognitive Skills that are Used for the Main Model and Their Measures from 1922 Questionnaire

Conscientiousness	Openness	Extraversion
Prudence and forethought	Desire to know	Fondness for large groups
Conscientiousness	Originality	Leadership
Truthfulness	Intelligence	Popularity with other children

Notes: Measures are named the same way as they appeared in the Terman questionnaire. See Figure A-1 of the Web Appendix for an example of Terman questionnaire question.

shows measures of Conscientiousness, Openness, and Extraversion, which were collected in 1922. As I am interested in the effect of childhood skills, measures from 1922 corresponding to ages around 12, are the most relevant. To complete the Big Five taxonomy, I also use measures of Neuroticism and Agreeableness from 1940 as the earliest available information on these two dimensions²¹. As I argue below, the 1940 factors of Agreeableness and Neuroticism show no association with longevity or education, either individually or jointly, and so they are used for robustness checks rather than for the main model estimation.

4 Methodology

4.1 A Model of Schooling and Longevity

I use the full maximum likelihood estimator and the expectation-maximization (EM) algorithm to estimate a system of equations that model relationships between latent noncognitive skills Θ , IQ , schooling D , and the hazard of death $\lambda(t)$.²² Estimation is conditional on picture of children’s functioning.

²¹See Web Appendix A for more details.

²²The IQ is not latent in this paper, as there is only one measure of IQ available, which is not enough for latent skill identification. However, this data limitation is minor, as the population is already selected on having extraordinarily high IQ s. Therefore, I concentrate on effects from personality, which has substantial variation, and use IQ mainly as an additional control just in case IQ still affects outcomes for this high-ability population.

background variables \mathbf{X} , which I omit from the equations for the sake of brevity.²³

Let D be a categorical choice of the highest education level obtained in life that takes values from 1 to 3: (1) high school graduate, (2) some college education, and (3) bachelor's degree or above.²⁴ Denote three binary variables for education levels as $D_k = 1[D = k]$, $k = 1, 2, 3$. In the ordered logit model, the latent propensity for education, D^* , determines the choice of schooling by the following rule: $D = 1$ if $D^* \leq \delta_1$, $D = 2$ if $\delta_1 < D^* \leq \delta_2$, and $D = 3$ if $D^* > \delta_2$, where δ_1 and δ_2 are thresholds to be estimated and ϵ_D is logistic distributed. D^* is affected by personality Θ and IQ:

$$D^* = a_1\Theta + a_2IQ + \mu_D + \epsilon_D, \quad (1)$$

where ϵ_D is an idiosyncratic error term and μ_D is unobserved heterogeneity in D^* .

A mixed proportional hazard (MPH) model with nonparametric baseline hazard function $\lambda_0(t)$ (like in the Cox model) relates the hazard of death after age 30, $\lambda(t)$, to noncognitive skills, IQ, and binary variables for education levels, D_1 and D_2 :

$$\lambda(t) = \lambda_0(t) \exp(\mathbf{b}_1\Theta + b_2IQ + b_3D_1 + b_4D_2 + \mu_\lambda), \quad (2)$$

where term μ_λ accounts for unobserved heterogeneity in $\lambda(t)$.²⁵

²³To impute missing values in \mathbf{X} , I use the MCMC multiple imputation method (Rubin, 1987; Schafer, 1997), which is known to preserve the variance-covariance matrix of data.

²⁴“Bachelor’s and above” category includes bachelor’s, master’s, and doctorate degrees. Since master’s and doctorate do not contribute to longevity of Terman subjects relative to bachelor’s, I do not separate them to keep the model as parsimonious as possible. The Wald test statistic for Masters and Doctorate jointly affecting longevity in the main model with added Masters and Doctorate levels is 0.461, with p -value of 0.794.

²⁵In general, coefficients of such model can be strongly biased due to dynamic selection problem. However, the age of the left truncation makes a difference for the practical importance of this bias. I find that model results are robust to the truncation point: truncation at start of the survey (year 1922, about age 12), and at ages 23, 25, 27, and 30 all show very similar results due to negligible mortality (see Table B-1 of the Web Appendix), which would not be the case if dynamic selection were a serious issue. These results are in line with Bijwaard et al. (2015), who note that “...education choice is determined before mortality plays a major role: mortality can be largely ignored for young ages.” As my model is robust to left truncation point before age 30, I use truncation at age 30 for the main model specification to avoid reverse causality, which happens automatically when students die before finishing their education. By age 30 only 5% of the Terman population are in school (see Figure B-1 of Web Appendix B).

To model unobserved heterogeneity μ_D and μ_λ I use the semi-parametric heterogeneity model approach, which has been shown to be effective for accurate estimation of both duration models (Heckman and Singer, 1984), and recursive nonlinear systems with endogenous variables (Mroz, 1999). This method is free from any specific functional form assumptions about the distribution of unobserved heterogeneity. Terms μ_D and μ_λ take on values $\mu_{D1}, \dots, \mu_{DK}$, and $\mu_{\lambda 1}, \dots, \mu_{\lambda K}$ with shared probabilities p_1, \dots, p_K , so that $\sum_{k=1}^K p_k = 1$. The number of latent classes K is determined by the econometrician to maximize statistical fit while avoiding empirical non-identification (e.g., Cameron and Trivedi, 2005, pp. 624–625).

I relate latent noncognitive skills to their noisy measures through measurement system (3), which is necessary for identification of the factor model. The measurement system, which allows for correlations among latent skills, can be described by the following equation:

$$M = \alpha\Theta + \eta, \quad (3)$$

where M is a vector of measures generated by a vector of latent factors Θ , α is a matrix of factor loadings that is empirically justified in Web Appendix A, and η is a vector of error terms.

I allow unobserved heterogeneity μ to be correlated with IQ and Θ : latent noncognitive skills and IQ are linked to probabilities p_1, \dots, p_K through a multinomial logit model:

$$p_k = \frac{\exp(c_{1k}\Theta + c_{2k}IQ)}{\sum_{j=1}^k \exp(c_{1j}\Theta + c_{2j}IQ)} \quad (4)$$

Using multinomial logit is a standard way to allow for correlation between unobserved heterogeneity and covariates (e.g., Huang and Bandeen-Roche (2004)). For background controls other than IQ I use the standard assumption of orthogonality between background controls and unobserved heterogeneity (Van den Berg, 2001). The likelihood function for the model is presented in Web Appendix C.

Identification of the model comes from well-established identification of its standard components. I make several standard normalizations and assumptions.²⁶ Those include:

²⁶See Anderson and Rubin (1956) for more details on standard factor modeling.

(a) setting the variance of each component of vector Θ to one; (b) setting the sign of the effect of each factor on its first measure so that the interpretation of the factor is positive²⁷; (c) assuming mutually independent error terms, η and ϵ_D ; (d) making a number of exclusion restrictions in matrix α , which are justified by exploratory and confirmatory factor analysis documented in Web Appendix A.

Second, the recursive nature of Equation (2) relative to the ordered logit model (1) also contributes to identification by providing a useful exclusion restriction: schooling decision appears in the right-hand side of the MPH model (2), but realized life duration, a future outcome, does not appear in the schooling model (1).²⁸ Our control for unobserved heterogeneity μ allows us to control of unobserved characteristics that predict both schooling and longevity. Subjects do not know exactly how long they are going to live when they make schooling decisions, but they may have expectations based on personal characteristics, such as general health. The schooling model is conditional on predictors of longevity including those unobserved to the econometrician (unobserved heterogeneity). Identification of the semiparametric heterogeneity model that allows for covariate effect on unobserved heterogeneity is shown by [Huang and Bandeen-Roche \(2004\)](#).

Under the assumptions of the model we can interpret results as causal. It is always safe to interpret results as associations conditional on a large number of observables and unobservables. Confounding is possible but given the quality and wealth of controls we can expect it to be minor. The first powerful control is selection of the sample on high IQ, which makes all subjects similar with regard to a known major confounder. The second is IQ itself, which is used as a regressor in case it still influences outcomes in such a selective sample. The third is a large set of observables including multiple early health measures, early educational investments by parents and private tutors, parental background, and cohort. The fourth is a set of well-established latent noncognitive skills represented by the Big Five taxonomy. The fifth is the control for permanent unobserved heterogeneity that is not accounted for by latent skills.²⁹ After taking full advantage of uniquely detailed data it

²⁷For instance, it is helpful when higher latent factor of Conscientiousness implies a higher, not lower, propensity to be organized and controlled.

²⁸See [Maddala \(1983\)](#), pp. 120–123 for an analysis of the useful role of recursion in identification of nonlinear systems of equations.

²⁹Examples of permanent unobserved heterogeneity include genetic endowments and

is hard to think of important confounders that remain uncontrolled.

The problem of reverse causality, which is estimating the effect of longevity on education instead of the effect of education on longevity, is unlikely. Individuals do not know their exact longevity when they make schooling decisions, but they may act on their longevity expectations. I do the following to account for such expectations: (1) controlling for several observed measures of early health of the individual and detailed family background; (2) using the sample conditional on survival to age 30, thus making the reverse causal case unlikely, since those who died as a student or soon after graduation are excluded³⁰; (3) finally, unobserved heterogeneity μ is designed to account for otherwise uncontrolled health-related shocks, attitudes, endowments, or family background.

4.2 Modeling Health-Related Outcomes and Life Satisfaction

To understand the mechanisms behind the effects of skills and education on longevity, I estimate parameters of the following equations for a number of health-related and life satisfaction outcomes H_p , $p = 1, \dots, P$, simultaneously with equations (1), (3), and (4):

$$H_p^* = d_1\Theta + d_2IQ + d_3D_1 + d_4D_2 + \mu_p + \xi_p, \quad (5)$$

where H_p^* is defined differently for binary and continuous outcomes. For binary outcomes, H_p^* is a latent propensity to make a choice in a logit model. For continuous outcomes, $H_p^* = H_p$ in a linear model. Letter ξ_p represents the error term. Other notation is the same as in Equation (2).

Health and health behaviors are likely mechanisms behind the effect of education on longevity and can, as a part of decomposition analysis, be added to the longevity equation 2. This, however, would change the interpretation of education coefficients. Instead of estimating the total effect of education, as I do now, I would estimate a direct effect conditional health shocks before making the choice of one's education level. I leave generalizations of this model that could include a dynamic modelling of health formation and the associated time-varying unobserved heterogeneity to future research based on data that are better fitted for such model.

³⁰By age 30, only about 5% of the Terman population are in school (see Figure B-1 of the Web Appendix).

on its mediators. This paper lacks both statistical power and data about several important mechanisms for a satisfactory decomposition analysis. For example, Terman data have no information about both early and mid-life smoking, an important mechanism (de Walque, 2010). Therefore, I limit my analysis to suggesting a number of observed mechanisms: I estimate how education affects variables that can be expected to affect longevity, such as heavy drinking. I leave full decomposition analysis to Hong, Savelyev, and Tan (2020), who use different data with a larger sample size and more detailed observed health behaviors.

4.3 Placebo Tests

Additional support of the model comes from placebo tests. Instead of outcomes, such as longevity or health at midlife, for which I find strong effects of education, for placebo tests I use eight early health measures, such as low birth weight. Obviously, college education cannot possibly affect early health, but if my results are driven by unobserved third factors, such as unobserved health endowments, it is possible to get statistically significant “effects” of education on early health, which would be an indication of a model that misses important confounders and only identifies associations. Estimating equation (5) simultaneously with equations (1), (3), and (4) and using early health outcomes as H_p , I find that none of the college education thresholds affects any of the eight measures of early health.³¹

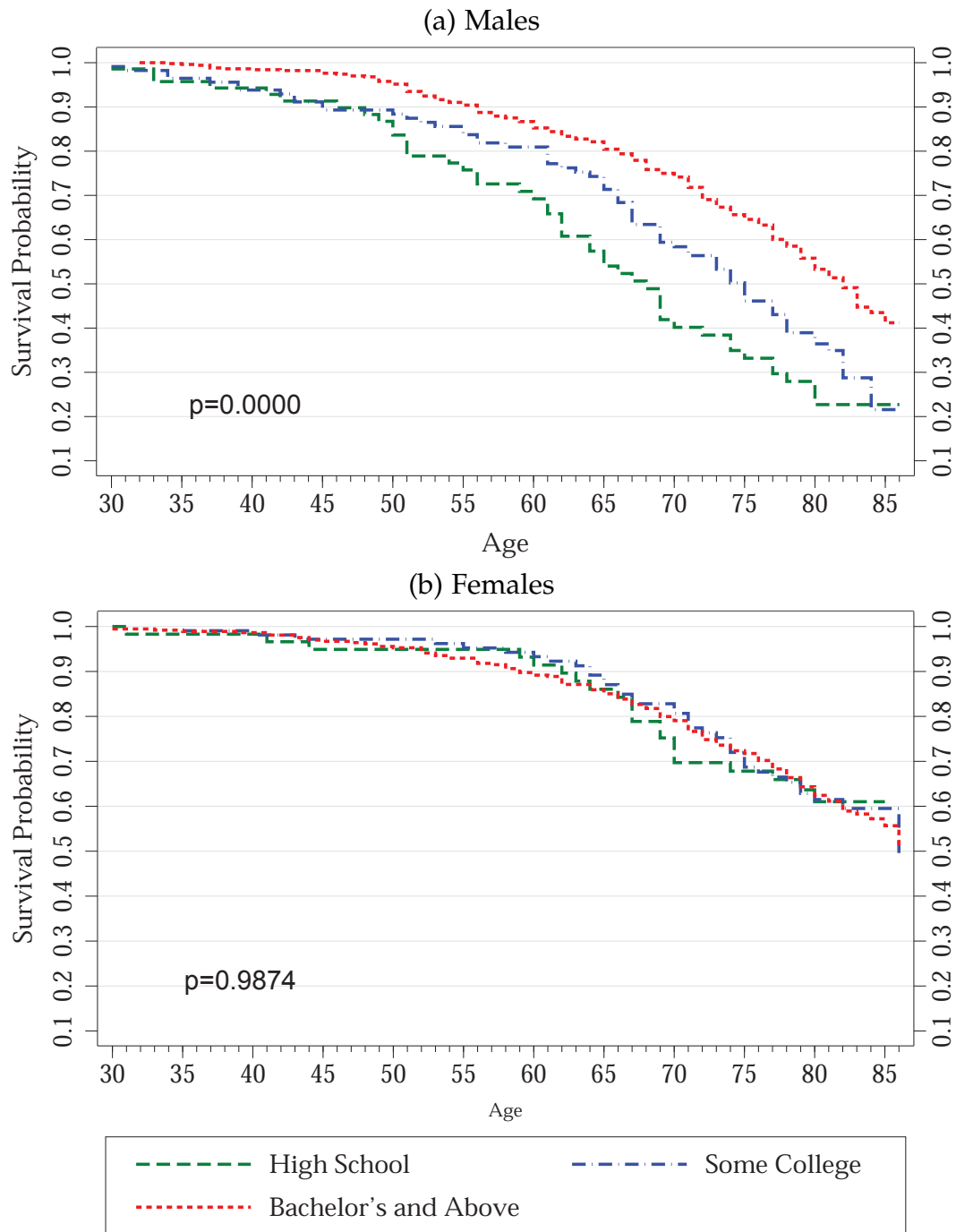
5 Results and Discussion

First, I describe nonparametric results and then proceed with the main (semi-parametric) model. Next, I discuss effects on life expectancy and the value of life, mechanisms behind the estimated effects, compression of morbidity hypothesis, robustness checks, limitations of results, and external validity.

Nonparametric Results Figure 1 shows the Kaplan-Meier survival curves by education and sex conditional on survival to age 30. For males, higher levels of education correspond to higher probability of survival. For instance, while for high school graduates the survival

³¹For placebo tests, the early health variable that is used as outcome is omitted from the set of controls X (see Table 8).

Figure 1: Kaplan-Meier Survival Function by Education



Notes: Probability of survival is conditional on survival to age 30. Education groups are mutually exclusive and refer to the highest level of education obtained in life; p -values are shown for the log-rank test for the equality of survivor functions. Calculations are based on the Terman data.

rate to age 80 is about 22%, for men with at least a four-year college degree the survival rate is 56%. The log-rank test for the equality of survivor curves is rejected.

For women, we can see no difference between the Kaplan-Meier survival curves by education level, and the log-rank test cannot be rejected. This result is preserved when I estimate the main model, which controls for both observed and unobserved heterogeneity. Cognitive and noncognitive skills show no effects for women either, either individually or jointly. These sex differences are consistent with results found by [Savelyev and Tan \(2019\)](#), who show that effects of education and noncognitive skills on health behaviors are substantially more numerous and stronger for men than for women of the high-IQ population. [Conti and Heckman \(2010\)](#) study the heterogeneity of the effect of education on health-related outcomes including smoking, poor health, and obesity, and find that the effect of education for men of the general population tends to become stronger with IQ, while for women it tends to become weaker. These tendencies are consistent with a strong effect for men and no effect for women for the Terman sample of high-ability individuals.

Model-Based Results by Sex As I find no relationships of interest for women, the rest of the paper focuses on results for men. Corresponding statistically insignificant main model results for women are presented in [Web Appendix B](#).

Main Model Estimates and the Role of Controls [Table 4](#) contains estimates of the main model for men (see [Model 1](#)), as well as estimates of simplified models (see [Models 2–5](#)).³² [Column \(1\)](#) shows coefficients of the main MPH model of mortality. Both incomplete college and high school education lead to higher mortality relative to completing a college degree. The estimates suggest that high school graduates are 2.6 ($= \exp(0.974)$) times more likely to die at any moment of time than those with a bachelor’s degree, and those with some college education are 1.7 ($= \exp(0.525)$) times more likely to die compared to college graduates.

We can also see that Conscientiousness and Extraversion show negative direct effects of similar magnitude on the hazard of death. As effects of Conscientiousness and Extraversion

³²See [Table B-2](#) of the [Web Appendix](#) for an analogue of [Table 4](#) estimated for females, which shows no statistically significant effects, neither individually nor jointly.

Table 4: Coefficients of MPH Model of Mortality and Ordered Logit Models of Education Choice, Males

	Model 1 (main model)		Model 2		Model 3		Model 4		Model 5		PH-test
	Mortality MPH (1)	Education Ordered logit (2)	Mortality MPH (3)	Education Ordered logit (4)	Mortality MPH (5)	Education Ordered logit (6)	Mortality MPH (7)	Mortality MPH (8)	Mortality MPH (8)		
Education											
High school degree	0.974 *** (0.215)	-	0.774 *** (0.188)	-	0.749 *** (0.191)	-	0.754 *** (0.190)	0.821 *** (0.175)	0.258		
Some college education	0.525 *** (0.167)	-	0.394 *** (0.144)	-	0.423 *** (0.138)	-	0.473 *** (0.137)	0.510 *** (0.134)	0.547		
Bachelor's degree or above	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted			
Skills											
Conscientiousness	-0.160 * (0.086)	0.192 (0.141)	-0.129 * (0.073)	0.210 * (0.127)	-	-	-	-	0.584		
Extraversion	-0.158 ** (0.076)	-0.195 (0.135)	-0.138 * (0.071)	-0.163 (0.125)	-	-	-	-	0.557		
Openness	0.095 (0.088)	0.155 (0.157)	0.093 (0.082)	0.178 (0.147)	-	-	-	-	0.547		
IQ	0.040 (0.069)	0.399 *** (0.135)	-0.025 (0.062)	0.299 *** (0.105)	-0.010 (0.061)	0.325 *** (0.099)	-	-	0.277		
Joint test <i>p</i> -values	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.727		
Unobs. Heterogeneity (μ)	Yes	No	No	No	No	No	No	No	No		
Background controls (X)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No		

Notes: Model coefficients are presented with standard errors shown in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Sample size is 695 for all models. Model 1 is the most preferred. Models 2–5 are simplified versions of Model 1 that are missing one or more types of controls. Estimates for the measurement system are presented and described in Web Appendix (see Table B-3). Corresponding estimates for females are statistically insignificant for all key factors both individually and jointly and are presented in Table B-2 of the Web Appendix. Results are based on the Terman data.

on education are at best weak, direct effects of these skills are close to their total effects.³³ I do not find a statistically significant effect of Openness, but positive sign of the estimate is consistent with adverse effects of Openness on health behaviors and life satisfaction, which is documented below.

Column (2) shows coefficients for the main ordered logit model of education. The only statistically significant result is a positive coefficient for IQ. It is remarkable that IQ shows a strong effect on education even though subjects were selected on IQ above 140. One reason for finding no statistically significant effect of noncognitive skills on education is that I control for unobserved heterogeneity. A comparison of columns (2) and (4) shows that without controlling for unobserved heterogeneity, the model shows a statistically significant relationship between Conscientiousness and education. This finding is likely driven by an unobserved confounding factor.

Another reason why controlling for unobserved heterogeneity is important is that estimated coefficients differ substantially with and without such a control. These results are in line with a well-known property of the MPH model that ignoring unobserved heterogeneity leads to a downward (by absolute value) bias of estimates (e.g., [Van den Berg, 2001](#)). Indeed, the average downward bias due to omitting μ in the MPH model is about 20% (compare statistically significant coefficients in columns (1) and (3)).

Models 3, 4, and 5 show results of incremental omission of variables from the model: noncognitive skills, cognitive skills, and background variables. Interestingly, as a result of successive changes from Model 1 to Model 5, coefficients decrease and then increase again and get closer to coefficients in the main model so that the coefficient for high school degree in Model 5 is biased 16% relative to that in Model 1, while the coefficient for some college is biased only by 3%. This peculiar result could not be foreseen prior to estimation and might be specific to this particular population.

Another unforeseen result of Table 4 is that the selection effect is at best small relative to selection found in the literature for the general population. The omission of all noncognitive skills changes estimated education coefficients by 5% on average.³⁴ The likely reason for

³³Additional evidence for this is available from the author upon request.

³⁴I use average absolute value of change, as coefficients change in different directions. The average value of change without taking the absolute value is 2%.

this result is either weak or nonexistent links between noncognitive skills and education for this population, as we see in Column (2). The omission of IQ, a single skill variable that does matter for education, changes education coefficients by 6% on average.

Table 4 also offers evidence that the proportional hazard assumption behind the main MPH model is valid. Column (9) shows p -values for the proportional hazard (PH) test. The PH hypothesis cannot be rejected for either individual variables or jointly.³⁵

Results of the MPH model of all-cause mortality are robust to the exclusion of individuals with specific major causes of death, as documented in Web Appendix B.³⁶ In another robustness check documented in Web Appendix B, I demonstrate that latent factors of Agreeableness and Neuroticism that are proxied by 1940 measures are not predictive of longevity both individually and jointly but greatly increase the degrees of freedom. These factors can be dropped for greater parsimony and to avoid mixing childhood and adulthood skills. The lack of association between these two factors and longevity is in line with previous research based on the same data (Martin et al., 2007). Also, the literature provides no strong prior for either positive or negative effects of these traits (see Chapman et al. (2011)).³⁷

Model-Estimated Survival Curves and a Comparison with the General Population Figure 2 shows model-predicted survival curves by education and compares them with a survival curve for white men of the general population, the 1910 birth cohort. First, predicted curves are very similar to non-parametric Kaplan-Meier curves in Panel (a) of Figure 1, which is another representation of the result that for this sample the combined selection effect on observables and unobservables is negligible.

Second, we can see a sizable differences in counterfactual survival by education. Conditional on survival to age 30, the chances to survive to age 86 is 40% for college graduates, 22% for those with some college education, and 10% for high school graduates.

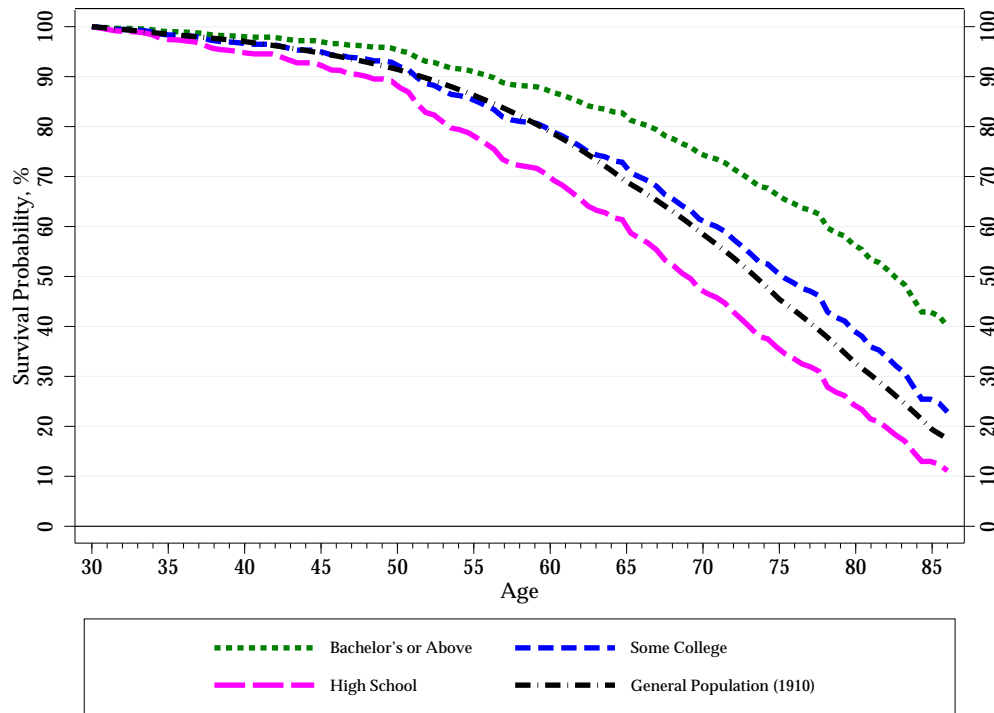
Finally, we can see that the general population of 1910 birth cohort of men has about 17% chance to survive to age 86, which is higher than for Terman high school graduates

³⁵For the PH test, I allow for time-dependence of coefficients and test the null hypothesis of no time-dependence.

³⁶See Table B-4.

³⁷See Table B-5.

Figure 2: Model Prediction of the Survival Function by Education and Comparison with General Population, Males



Notes: Calculations are based on the Terman data and Census for Disease Control (CDC) (Arias, 2012) for the survival of the 1910 cohort of white men.

but lower than for Terman subjects with some college education. The result that Terman subjects with a high school degree live shorter than men from the general population, despite having superior intelligence and education, is a puzzle. I address this puzzle below in a section on the mechanisms and provide evidence that this result is consistent with life dissatisfaction and adverse health behaviors of Terman high school graduates likely arising from a major mismatch between their extraordinary cognitive ability and types of careers available to high school graduates.

Effects of Skills on the Hazard of Death I find that both Conscientiousness and Extraversion have statistically significant total effects on longevity and that these effects are not driven by the choice of education, as we can foresee from panel 2 of Table 4.

A non-negative association between Conscientiousness and longevity is consistent with a substantial literature (see [Kern and Friedman \(2008\)](#) for a survey). In particular, [Friedman et al. \(1993\)](#) found a beneficial association for high-ability men based on the same data. The authors used the Cox PH model with 1–7 degrees of freedom. I show that this qualitative result survives controlling for detailed observed and unobserved heterogeneity. My total effect estimate implies that an increase in Conscientiousness by one standard deviation (s.d.) leads to a 17% decline ($(1 - \exp(-0.182)) \cdot 100\%$) in the hazard of death, a result that is statistically significant at the 5% level.³⁸

While Extraversion generally shows mixed associations with longevity, studies that use Sociability and related facets of Extraversion, as in this paper, find positive associations with longevity (see [Chapman et al. \(2011\)](#) for a survey). Previous studies based on the Terman data did not find any associations between Extraversion/Sociability and longevity (e.g., [Friedman et al., 1993](#)), but they did not account for IQ, background controls, unobserved heterogeneity, or measurement error in measures of personality. Thus, the strong effect of Extraversion on longevity that I find is a novel result for a high-ability population and a contribution to the ongoing debate about the role of Extraversion in longevity production. One standard deviation increase in Extraversion leads to a 13% decline ($(1 - \exp(-0.135)) \cdot 100\%$) in the hazard of death, a result that is statistically significant at the 10% level.³⁹

I find no statistically significant effects of Openness and IQ on longevity. As mentioned, I also find no evidence of longevity effects of Agreeableness and Neuroticism.

Effects of Skills and Education on Life Expectancy Column (1) of Table 5 shows counterfactual life expectancies at age 30 by education level. It also shows that the model-predicted sample average is close to the sample average implied by nonparametric Kaplan-Meier estimates. Column (2) shows differences between counterfactual life expectancies by education level relative to high school education. Conditional on survival to age 30, some college education adds 4.7 additional years of life, while a 4-year college degree or above adds 9.8

³⁸My estimate of the total effect of Conscientiousness on the hazard of death is -0.182, with a standard error of 0.086. Results on the decomposition of the total effect into direct and indirect effects were not included into the final version of the paper, but they are available from the author upon request.

³⁹My estimate of the total effect of Extraversion on the hazard of death is -0.135 with a standard error of 0.077. More details are available from the author upon request.

Table 5: Effects of Education on Life Expectancy at Age 30, the Corresponding Monetary Value for a Statistical Person, and Robustness Checks, Males

	Method of extrapolation of the baseline survival function						
	CDC data (main method)		Linear		Gompertz-Makeham		Standard deviation of alternative estimates ^(c)
	Estimated parameter ^(a)	Total effect ^(b)	Estimated parameter ^(a)	Total effect ^(b)	Estimated parameter ^(a)	Total effect ^(b)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
A. Life expectancy at age 30, years							
High school	38.7	-	39.2	-	38.9	-	-
Some college	43.4	4.7	44.0	4.9	43.9	5.0	0.14
Bachelor's degree or above	48.4	9.8	48.9	9.7	49.4	10.5	0.46
Conscientiousness	48.6	2.0	49.1	2.0	49.7	2.2	0.14
Extraversion	48.2	1.6	48.7	1.7	49.2	1.8	0.11
Sample average, model	46.6	-	47.1	-	47.5	-	-
Sample average, Kaplan-Meier	46.2	-	48.7	-	48.2	-	-
B. Value of remaining life, mln. 2016 USD							
High school	9.5	-	9.5	-	9.5	-	-
Some college	10.0	0.50	10.0	0.49	10.0	0.50	0.0017
Bachelor's degree or above	10.4	0.93	10.4	0.92	10.4	0.93	0.0078
Conscientiousness	10.4	0.16	10.4	0.15	10.4	0.16	0.0011
Extraversion	10.3	0.11	10.3	0.11	10.3	0.11	0.0005
Sample average, model	10.2	-	10.2	-	10.2	-	-
Sample average, Kaplan-Meier	10.2	-	10.2	-	10.2	-	-

Notes: ^(a)Counterfactual life expectancy is calculated for (1) different levels of education given the average levels of skills and background variables and (2) skills higher than average by one standard deviation accounting for both direct and indirect effects though education. ^(b)The education effect is calculated relative to the baseline level of longevity of high school graduates. For skills, effects are relative to the average level of skills. ^(c)Standard deviation represents variation created by three alternative methods of interpolation. Dashes represent non-applicable cases. Calculations are based on the Terman data.

additional years of life, a remarkably high benefit.⁴⁰ The same column shows that changing Conscientiousness and Extraversion by one standard deviation leads to 2.0 and 1.6 additional expected years of life respectively.

The baseline survival function is nonparametrically estimated as part of the MPH model until age 86, which allows estimation of the main model survival function until that age. However, for a life expectancy calculation I need full support of the longevity distribution. To solve this problem, I extrapolate the baseline survival function to age 100 using census data for men of the cohort born in 1910. As a robustness check, I compare this extrapolation with a linear extrapolation. As the upper tail of the survival curve is expected to be convex, the linear extrapolation is, arguably, an upper bound. I also use a Gompertz-Makeham extrapolation (e.g., [Brown et al., 2002](#)), an alternative approach to using Census survival rates. Results based on these three methods are similar. The standard deviation of estimates based on these three methods of extrapolation is shown in Column (7) and can serve as an approximate estimate of the error introduced by the extrapolation. For the effect of bachelor's degree (9.8 years), two standard deviations give us a confidence width of about one year.⁴¹

Effects on Skills and Education on the Value of Remaining Life In Panel B of Table 5, longevity benefits are evaluated in 2016 USD for a statistical person, using Murphy and Topel's (2006) methodology. The present value of additional life expectancy at age 30 compared to high school graduates is \$500,000 for some college education and \$930,000 for 4 years of college. These numbers represent the value of increase in life expectancy induced by education for a contemporary 30-year old man if effects estimated in this paper were applicable to him. This high longevity gain is estimated as a lower bound, as it does not include the value of increased quality of life in the form of superior general and mental health over the lifespan.⁴² The valuation of one standard deviation of Conscientiousness and Extraversion are \$160,000 and \$110,000 respectively. Thus, the effects of education and

⁴⁰I found no evidence of additional effects on longevity from master's and doctorate relative to bachelor's.

⁴¹Before age 86 the MPH model coefficients are time-invariant, as supported by the PH test. The implicit assumption that coefficients stay the same after age 86 may introduce an additional error.

⁴²I discuss the effect on general health below, as part of a discussion of Table 6.

skills are economically significant. Unlike Panel A, Panel B shows almost identical results for all three methods of extrapolation. The robustness of the value of remaining life comes from discounting the future: the extrapolation only affects ages above 86, which are heavily discounted at age 30.⁴³

Effects of Education and Skills on Health and Health-Related Outcomes Table 6 shows effects of education and skills on health and health-related outcomes. Models presented in the table are similar to the longevity model discussed above. Just as for longevity, model (5) for behavioral and health outcomes is estimated jointly with the models (1), (3), and (4) conditional on the same observable and latent controls. Estimation is performed outcome by outcome.

The observed effect of education on longevity is supported by several plausible mechanisms documented in Table 6. Indeed, we observe statistically significant positive effects of education on marriage, memberships in organizations, occupation as a professional, and general health, and negative effects on heavy drinking of alcohol and mental difficulty.

Conscientiousness is known to have a strong association with beneficial health-related outcomes (Roberts et al., 2007). Conscientious people tend to delay gratification, plan for the future, and act towards their goals (John and Srivastava, 1999). These characteristics boost health-beneficial choices. Not surprisingly, Table 6 shows that conscientious subjects are less likely to drink heavily, more likely to get married, and have better general and mental health. These results suggest mechanisms behind the relationship between Conscientiousness and longevity and support a causal interpretation.

The effect of Extraversion on longevity is also supported both theoretically and empirically. Extraversion, which is a propensity to be active and social, may help create social skills and networks of friends, which, in turn, boost both mental health and earnings.⁴⁴ Gensowski (2018) shows a strong effect of Extraversion on earnings with the same data. However, it is unclear from the literature whether earnings (or, in general, wealth) affect

⁴³I use discount rate 3.5 following Murphy and Topel (2006).

⁴⁴Extraverted individuals tend to have larger social networks and more friendships, however the quality of these friendships is an ongoing area of research (see Harris and Vazire (2016) for a survey).

Table 6: Effects of Education and Personality on Health-Related Life Outcomes, Males

	Ever involved in heavy drinking in 1940–1960 (about ages 30–50)	Married once and still married in 1960 (about age 50)	Number of memberships in 1960 (about age 50)	Belongs to professional group in 1959 (about age 49)	Mental difficulty in 1950 (about age 40)	Mental difficulty in 1960 (about age 50)	General health rating in 1950 (about age 40) ^(a)	General health rating in 1972 (about age 62) ^(a)	General health rating in 1986 (about age 76) ^(a)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Marginal Effect of Education									
High school	0.134 * (0.076)	-0.135 * (0.074)	-1.205 *** (0.313)	-0.519 *** (0.108)	0.133 ** (0.065)	0.070 (0.082)	-0.234 * (0.171)	-0.64 *** (0.275)	-0.704 *** (0.311)
Some college	0.092 (0.061)	-0.205 *** (0.056)	-1.549 *** (0.258)	-0.477 *** (0.077)	0.042 (0.051)	-0.003 (0.062)	-0.296 ** (0.129)	-0.421 *** (0.177)	-0.13 (0.228)
College or above	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Total Marginal Effects of Skills^(b)									
Conscientiousness	-0.077 *** (0.028)	0.055 ** (0.028)	0.089 (0.136)	0.004 (0.029)	-0.073 *** (0.026)	-0.071 ** (0.029)	0.141 *** (0.054)	0.121 * (0.066)	0.074 (0.086)
Openness	0.051 * (0.031)	-0.067 ** (0.030)	0.068 (0.148)	-0.034 (0.029)	0.097 *** (0.028)	0.102 *** (0.033)	-0.252 *** (0.060)	-0.117 (0.073)	-0.027 (0.084)
Extraversion	0.06 ** (0.029)	0.044 * (0.026)	-0.017 (0.130)	-0.022 (0.027)	-0.076 *** (0.024)	-0.085 *** (0.027)	0.143 ** (0.056)	0.139 ** (0.056)	0.162 *** (0.077)
IQ	0.028 (0.028)	-0.020 (0.025)	0.042 (0.121)	0.025 (0.027)	0.005 (0.021)	0.022 (0.024)	-0.068 (0.056)	-0.081 (0.067)	0.028 (0.094)
Joint test <i>p</i> -value: ^(c)	0.0001	0.0022	0.0000	0.0000	0.0024	0.0137	0.0010	0.0017	0.0351

Notes: Estimates are based on jointly estimated system of equations (1,3,4,5). ^(a)One-tailed tests are used for education coefficients affecting health based on consistent evidence from the literature on the nonnegative effect of education on health (Galama et al., 2018; Grossman, 2015; Lochner, 2011). ^(b)Total marginal effect is the sum of direct and indirect (though education) marginal effects. ^(c)Joint test is performed for all variables listed in this table. All estimation is conditional on background variables from Table 1 and unobserved heterogeneity. Standard errors are in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Estimation sample size is 695. Results are based on the Terman data.

health and longevity and, if yes, whether the effect is positive, negative, or nonexistent.⁴⁵

As for other mechanisms, Table 6 shows that Extraversion is associated with one adverse health behavior, heavy drinking, which might be a side effect of greater socialization (Panel 1). However, positive effects on general and mental health suggest that negative contribution of heavy drinking is compensated by other channels. I document a positive effect of one such possible channel, stable marriage (Panel 2).

As is also discussed in a companion paper on health behaviors by [Savelyev and Tan \(2019\)](#), Openness for high-ability men shows no single beneficial health-related outcome and a number of adverse ones. In particular, we see from Table 6 that heavy drinking increases, marriage probability decreases, and mental and general health decrease. However, the effect on general health fades after age 40, which might be the reason why we do not observe any statistically significant adverse effect of Openness on longevity.

Compression of Morbidity The beneficial effects of education, Conscientiousness, and Extraversion not only on longevity but also on health over the lifecycle found in this paper are consistent with the controversial compression of morbidity hypothesis: whether the fraction of the lifespan spent in good health increases with longevity (e.g., [Cutler et al., 2014](#)). Compression of morbidity has important practical implications. First, it implies that the highly valuable superior *quantity of life* is accompanied by greater *quality of life*, thus making increases in longevity even more beneficial for individuals ([Murphy and Topel, 2006](#)). Therefore, an increase in longevity through investments in skills and education does not create an army of disabled people, but people who are both healthier and live longer. Finally, the compression of morbidity lowers the cost of health care relative to the case of no compression.

Results of this paper are consistent with the compression of morbidity. As we see from Table 6, the effect of schooling and Extraversion on general health are economically

⁴⁵[Snyder and Evans \(2006\)](#) find a negative effect of wealth on longevity and [Kippersluis and Galama \(2014\)](#) find a positive effect on adverse health behaviors. In contrast, [Schwandt \(2018\)](#) find a positive effect on health and longevity. [Frijters et al. \(2005\)](#), [Lindahl \(2005\)](#), and [Gardner and Oswald \(2007\)](#), find a positive effect on health related outcomes. Finally, [Kim and Ruhm \(2012\)](#), [Apouey and Clark \(2015\)](#), and [Cesarini et al. \(2016\)](#) find no effect on health related outcomes with the exception of some small reduction in the use of mental health drugs ([Cesarini et al., 2016](#)).

and statistically significant and persistent for ages 40–76.⁴⁶ Conscientiousness increases general health at ages 40 and 62, an effect that declines with age and becomes statistically insignificant by age 76. Schooling increases mental health at age 40. Conscientiousness and Extraversion increase mental health at both ages 40 and 50. Thus greater longevity that is induced by growing education and noncognitive skills is associated with superior general and mental health in both middle- and advanced ages.

Similar health-related outcomes are also analysed in a companion paper by [Savelyev and Tan \(2019\)](#), which is based on the same Terman data but does not study longevity. Apart from other contributions relative to earlier works by psychologists (e.g., [Friedman et al., 1995](#); [Martin et al., 2007](#)), [Savelyev and Tan \(2019\)](#) demonstrate that the key conditional associations between skills, education, and health-related outcomes survive controlling for familywise error, an error that originates from testing multiple hypotheses. In this paper, the aim of modeling health and health behaviors is different: I support the compression of morbidity hypothesis and provide evidence on the mechanisms using models that are fully consistent with the main longevity model.⁴⁷ In particular, Table 6 demonstrates results conditional on unobserved heterogeneity μ , which was not controlled for in previous literature (e.g., [Friedman et al., 1995](#); [Savelyev and Tan, 2019](#)).

Effects of Education and Skills on Life Satisfaction Table 7 is based on the same model as Table 6 but with different outcomes. It shows effects of education and personality on life and job satisfaction. Panels (1-8) suggest that lower level of education, especially high school degree, leads to lower probability of job and life satisfaction, lower control over choice of occupation, sense of wasted intellectual abilities, and hindered life due to inadequate education.

Another result of this table is that Conscientiousness is associated with greater work satisfaction, availability of outlets for mental capabilities, and job satisfaction, whereas Openness has negative effects on the very same outcomes. Combined with results of Table 6,

⁴⁶For instance, as late as at age 76, not having a high school degree increases health disadvantage relative to college graduates by 0.7 standard deviations (SD) of the general health distribution, a large change. One SD of Extraversion increases health by 0.16 SD, a smaller but not negligible change.

⁴⁷The models have the same education sub-model and measurement system, the same set of controls, and all account for unobserved heterogeneity.

Table 7: Effects of Education and Personality on Life Satisfaction, Males

	Definitely chosen life's work (1940)	Occupation chosen rather than drifted into (1940)	Work provides great satisfaction (1950)	Inadequate education hindered life (1950)	Life offers outlets for mental capabilities (1950)	Lived up to intellectual abilities (1950)	Good feelings about present vocation (1950)	Was early occupation choice fulfilled (1950)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Marginal Effect of Education								
High school	-0.171 ** (0.075)	-0.274 *** (0.081)	-0.167 ** 0.068	0.241 ** 0.117	-0.089 * (0.057)	-0.364 ** (0.189)	-0.393 ** (0.194)	-0.364 *** (0.112)
Some college	-0.050 (0.060)	-0.205 *** (0.058)	-0.08 0.054	0.174 ** 0.08	-0.092 ** (0.043)	-0.174 * (0.114)	-0.076 (0.138)	-0.248 *** (0.086)
College or above	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Total Marginal Effects of Skills^(a)								
Conscientiousness	0.020 (0.026)	0.032 (0.030)	0.05 ** 0.023	-0.015 (0.016)	0.04 ** (0.019)	0.166 *** (0.053)	0.14 ** (0.057)	0.045 (0.034)
Openness	-0.015 (0.029)	0.032 (0.030)	-0.089 *** 0.029	0.011 (0.020)	-0.075 *** (0.025)	-0.122 ** (0.062)	-0.147 ** (0.067)	-0.073 * (0.038)
Extraversion	-0.010 (0.026)	0.032 (0.030)	0.001 0.023	0.034 (0.023)	0.058 ** (0.024)	0.013 (0.061)	0.047 (0.066)	0.043 (0.035)
IQ	-0.041 (0.025)	0.032 (0.030)	0.007 0.024	0.021 (0.013)	-0.008 (0.021)	-0.081 (0.055)	-0.055 (0.057)	0.037 (0.036)
Joint test <i>p</i> -value: ^(b)	0.0022	0.0000	0.0024	0.0137	0.0010	0.0351	0.0351	0.0351

Notes: Estimates are based on jointly estimated system of equations (1,3,4,5). ^(a)Total marginal effect is the sum of direct and indirect (through education) marginal effects. ^(b)Joint test is performed for all variables listed in this table. All estimation is conditional on background variables from Table 1 and unobserved heterogeneity. Standard errors are in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Estimation sample size is 695. Results are based on the Terman data.

which showed increased heavy drinking combined with negative effects on marriage, mental, and general health, we can conclude that Openness is counterproductive for this cohort with respect to multiple observable health-related outcomes. In line with these results, the estimate of the total effect of Openness on the hazard of death is positive, but not precisely determined.⁴⁸

Understating Sizable Effects of Education on the Hazard of Death and Life Expectancy

The first question is whether estimated sizable effects represent causal relationships. As I argue in Section 4, while confounding and reverse causality cannot be completely ruled out, for this particular data they are unlikely to drive the results.

The second question is how to explain sizable effects of education for men: 4.7 additional years for some college education and 9.8 years for Bachelor's or above conditional on living to age 30.

The estimated effect is comparable with the effect found by [Lleras-Muney \(2005\)](#), who uses US compulsory schooling laws as a source of identifying variation and finds that an additional year of compulsory schooling causes a 1.7 year increase in life expectancy conditional on living to age 35. Conditional on a strong assumption of constant effect of schooling over different educational thresholds, this would imply about seven years for an additional four years of education, which is the difference between 4-year college and high school. Since we find no effect for women, our effect of college for a person of random sex is about five years ($0.5 \cdot 9.8 + 0.5 \cdot 0$). This is less than the seven years implied by the result by Lleras-Muney, which is based on a sample of pooled sexes. As mentioned, it is a strong assumption that the effect of compulsory schooling for a marginal school dropout is the same as the average effect of college education for a person with IQ above 140, but it is of interest that our complementing results are of comparable magnitude.

[Darden et al. \(2018\)](#) show that the longevity cost of lifetime smoking is about 4.3 years, about a half of our estimated effect of education. However, we can expect education of people with high ability to affect longevity not only through smoking but also through many

⁴⁸My total effect estimate implies that an increase in Openness by one s.d. leads to an imprecisely determined 8% increase in the hazard of death ($(\exp(0.077)-1) \cdot 100\%$). Coefficient is 0.077 and standard error is 0.089.

other mechanisms of longevity (e.g., [Savelyev and Tan, 2019](#)). A number of suggested mechanisms are documented in in this paper: reduced heavy drinking, stable marriage, increased social interactions, higher likelihood of having a professional occupation, superior mental and general health, and multiple measures of superior job and life satisfaction. As I argue below, job and life satisfaction may play a special role for this unusual cohort.

Two traditional explanations of the effect of education on health are (1) productive and (2) allocative efficiencies. Productive efficiency refers to larger health output from given amounts of endogenous (choice) inputs ([Grossman, 2000](#)). Allocative efficiency refers to better allocation of investments due to better information about the true effects of health investments ([Kenkel, 2000](#)). [Cutler and Lleras-Muney \(2008\)](#) offer another classification of mechanisms: (i) higher income; (ii) safer jobs; (iii) higher value of life; (iv) better health knowledge and superior cognitive skills; (v) lower discount rate and increased risk aversion (vi) higher rank in the society; and (vii) larger social networks, which provide financial, physical and emotional support. Sociologists ([Ross and Wu, 1995](#)) suggest three categories of explanations: (A) work and economic conditions; (B) social-psychological resources; and (C) health lifestyle.

The mechanisms mentioned above (1, 2, i–vii, A–C) cover many possible channels and partially overlap. The Terman sample provides evidence consistent with the following channels among those named above: higher income/economic conditions ([Gensowski, 2018](#)), higher rank in the society (measured by professional occupation), higher value of life (measured by greater life satisfaction), larger social networks/social-psychological resources (measured by number of memberships in organizations), and healthier lifestyles (measured by no heavy drinking and stable marriage). However, the channel though safer jobs is unlikely for this sample.⁴⁹

One channel related to effects of education on life and job satisfaction documented in

⁴⁹Occupational data suggest that the longevity gap is not explained by dangerous occupations such as mining, fishing, or cutting trees: no such occupations were reported. This is not surprising, as high school graduates were well-educated by standards of their time. In particular, in 1960, among high school graduates only five subjects worked in occupations that are associated with increased hazard of death or injury on the job: police, fire protection, and the military. Among college dropouts only three were with police or fire departments. All others had jobs with minimal mortality risks: clerical, sales, etc. Among college graduates none was with fire or the police, but four were with the military.

Table 7 deserves special attention for Terman subjects. This mechanism is separate from productive and allocative efficiencies and can be viewed as a special case of the effect through the value of life. I call this suggested mechanism *human potential-education matching*. One role that college education plays is that it helps select individuals with high potential, such as high IQ, and provides them with complementary skills that open doors to markets for high-skilled jobs. These high-skilled jobs provide a level of challenge that is a good match for people with high potential. There is anecdotal evidence that people who failed to realize their high potential may end up abusing substance or otherwise neglecting their health and living shorter lives. I provide statistical evidence consistent with this intuition, suggesting a channel through which IQ can increase the effect of education on health. So far, the literature suggested the opposite effect due to substitution of IQ for education in health production (Auld and Sidhu, 2005).

The Terman data are consistent with the hypothesis that a mismatch between education and IQ decreases health and longevity. First, the survival curve of Terman high school graduates is below the survival curve of the general population of US white men of the same cohort (see Figure 2), even though an average white man from this cohort had both lower intelligence and lower education than Terman's high school graduates.⁵⁰ I suggest an explanation for this paradox: Terman's high school graduates had a larger mismatch between human potential and their education than an average man from the general population, which likely resulted in less self-care for their health.

Second, Table 7 shows effects of high school and some college education relative to completed 4-year college degree on job and life dissatisfaction. We can see that lower education levels lead to a dislike of current occupation (Panels 1, 3, 7, and 8), lack of control in the choice of work (Panel 2), lack of intellectual challenge (Panels 5 and 6), and hindered life due to inadequate education (Panel 4). All these effects take place for those with high school degree and the majority of them also hold for those with an incomplete college education. These satisfaction results are consistent with effects of education on life outcomes from Table 6: a smaller likelihood of belonging to professionals (Panel 4),

⁵⁰White men in the general population have IQ of 100, as opposed to $IQ > 140$ of the Terman subjects. Only about 5% of the 1910 cohort finished college and only about 30% finished high school according to Figure 5 of Heckman and LaFontaine (2010).

a higher likelihood of heavy drinking (Panel 1), and greater difficulties with mental and general health (Panels 5–9)

Sociologists have investigated the hypothesis that education improves health through work satisfaction, suggesting that high-skilled jobs tend to be more rewarding and provide superior control over an employee’s life. However, the results of this literature are inconclusive, and intrinsic work rewards among the employed are found to be small, inconsistent, and not always positive (Ross and Wu, 1995). Perhaps, one reason behind this controversy could be that it is good matching between education and personal potential that matters rather than the level of education alone. Education that overmatches personal potential could be counterproductive for life satisfaction: the person could become overeducated for a number of jobs that could otherwise be a good fit, but still not a good match for challenging high-level jobs. In addition, jobs that require more education do not necessarily provide more control over a person’s life. For instance, a self-employed contractor with a high school degree can be expected to experience superior control over his life than a low-level office worker with a college degree. However, in the case of Terman high school graduates, the mismatch between intellectual ability and education level is major and one-directional, and the evidence for the job dissatisfaction effect is major.

Placebo Tests As part of the specification testing of my model, I run placebo tests for eight early health outcomes. I use the same model (1,3,4,5) as I use for health-outcomes in mid-life. This model has the same structure as the main longevity model of this paper except that instead of the hazard of death I use early health variables as outcomes one by one. Despite no control for familywise error rate, none of the p -values are statistically significant even at the 10% level (see Table 8).

Robustness to a Two-Step Estimation One common concern about using a full maximum likelihood model in research involving latent skills is that, in the case of model misspecification, latent skills might be primarily driven by outcomes rather than the measurement system. I show in Web Appendix B that this concern is not pertinent to results of this paper. Following Heckman et al. (2018), I estimate only the measurement system in the first step and then estimate the maximum likelihood model while fixing factor loadings of the mea-

Table 8: Placebo Tests

	High School			Some College		
	coefficient	standard error	<i>p</i> -value	coefficient	standard error	<i>p</i> -value
Normal Birth	0.049	0.066	0.459	0.053	0.052	0.308
No Breastfeeding	-0.033	0.037	0.371	0.051	0.039	0.191
Health Rating in 1922	0.123	0.127	0.333	-0.007	0.095	0.944
Low Birthweight	0.009	0.021	0.678	0.003	0.018	0.879
Mother's poor health during pregnancy	0.035	0.049	0.477	0.013	0.042	0.768
Nutrition Poor or Fair in 1922	0.001	0.035	0.967	0.047	0.033	0.154
Severe acute and chronic illness since 1922	-0.022	0.045	0.630	0.008	0.044	0.849
Signs of Nervousness	0.007	0.062	0.903	-0.072	0.049	0.144

Notes: Tests are based on the model of health outcomes (1,3,4,5) applied to early health measures as if they were outcomes. For each tested early health measure, the set of controls X excludes that tested measure.

surement system to the values estimated in the first step. The results are almost identical to those based on full maximum likelihood.⁵¹ With this concern shown to be alleviated, full maximum likelihood is a preferred estimator due to its superior efficiency and the lack of a need to correct standard errors for a two-step estimation.

Data Limitations I use a unique and detailed dataset that has early measures of skills and a long follow-up, but it is not without disadvantages. One disadvantage is its modest sample size. In addition, a number of important potential mechanisms, such as smoking and exercise, have not been measured in early life or in mid-life.⁵² Because of the modest sample size and missing data on potentially important mechanisms, I lack both statistical power and key information for decomposing the estimated total effects on longevity to a sum of underlying mechanisms. As a result, in this paper I only provide suggestive

⁵¹See Table B-6 of the Web Appendix.

⁵²Quitting smoking based on information from the 1964 Surgeon General Report might contribute to the total effect of education (de Walque, 2007). The identification of our treatment effect, however, does not rely on identification of mechanisms behind it.

evidence based on Table 6 and 7 and leave decomposition analysis for a separate paper based on a different dataset (Hong, Savelyev, and Tan, 2020).

Another limitation is the lack of measurement of childhood Agreeableness and Neuroticism. However, I show that Agreeableness and Neuroticism measured in adulthood do not predict longevity, suggesting that this disadvantage is unlikely to be crucial, which is in line with the literature.

Finally, having a censoring point for mortality in 1991 is another disadvantage. I account for the right censoring in the model, leading to internally-consistent estimates before the censoring point. However calculations of life expectancy and the value of remaining life required extrapolating the baseline survival function to more advanced ages. As shown, additional error introduced by this extrapolation is estimated to be small. A likely reason for an accurate extrapolation is that it starts from an advanced age of 86.

External Validity The tradeoff of analyzing quality longitudinal mortality data is the need to consider whether results are generalizable to a more contemporary population. I argue that for high ability males we can expect similar qualitative results for later cohorts. Indeed, we know from the literature that education and skills still determine lifestyles for the current population, which, in turn, still determine their longevity. Quantitatively, however, effects might be different for many reasons including significant progress in health knowledge and public health improvements over the course of the 20th Century.

Contemporary cohorts have better knowledge of the importance of health behaviors such as smoking, maintaining a healthy diet, and exercise. For instance, while students today learn about the adverse effects of smoking at school, earlier cohorts lacked widely-available consensus knowledge about smoking before the Surgeon General Report (Terry et al., 1964), at which point the Terman cohort was, on average, 54 years old.

Despite the lack of consensus knowledge about the dangers of smoking before 1950s (de Walque, 2010), substantial health information was widely distributed and available in 1922, the year when the survey began and respondents were about 12 years old. For example, "Rules of the Health Game," a public health pamphlet that was published in 1922 and targeted children, included personal hygiene, long sleep, eating vegetables and fruits, drinking at least four glasses of water a day, and playing part of every day outdoors

([Allensworth et al., 1997](#)). Alcohol was not mentioned, but given that Prohibition took place from 1920 to 1933, the adverse effects of alcohol were well-publicized at the time. Taking other drugs and obesity-related behaviors were not mentioned, but those were much less prevalent for this cohort.

The direction of quantitative change of the effect for later cohorts is unclear. On the one hand, since educated and conscientious people act on their health knowledge, the effects of education and Conscientiousness might be even stronger for later cohorts than what is estimated for the Terman cohort, as health knowledge grows over time. Larger incidence of obesity and the use of narcotics for current cohorts could increase the effect. On the other hand, smaller incidence of smoking for current cohorts may diminish the effect.

For females, the results should be considered historical since women's emancipation over the 20th century led to a considerably wider variety of lifestyles and health behaviors for women than before. Labor markets and the purpose of education also drastically changed for women. A wider variety of lifestyles together with greater health knowledge and labor force participation could create health effects of education and skills even if they were not strong or existent for earlier cohorts of women.

Presently, intelligent people tend to finish college in larger numbers than before. However, low-income students with high human potential are less likely to enroll in college than their higher income counterparts. Even after being accepted into college, low-income students are much more likely to choose not to attend ([Castleman and Page, 2014](#)). Moreover, college dropout rates substantially vary by socioeconomic status ([Bowen et al., 2009](#)). My paper shows that it is wrong to assume that high cognition substitutes for education in producing health and longevity. Moreover, I argue that the IQ may even increase the effect of education on health-related outcomes. This makes mismatches between human potential and education even worse than we may think from lost productivity considerations alone.

Effects of this paper might be similar to those for a larger population of able people, not only to extraordinarily able people. Indeed, understanding the best inputs into health production, such as a healthy diet, definitely requires cognitive skills, but not necessarily the abilities of a genius. The direction of effect change once the IQ gets smaller is unclear though: results by [Auld and Sidhu \(2005\)](#) suggest larger effects for lower IQs. However, results of [Conti and Heckman \(2010\)](#) for men and the human potential-education matching

mechanism suggested in this paper allow for a decline in the effect of education with a decline of the IQ.

6 Conclusions

This paper accounts for latent noncognitive skills developed in early life and unobserved heterogeneity in order to investigate relationships between skills, education, and longevity. The effect of education on health and longevity has implications for positive education subsidies in cases where education investments are at sub-optimal levels. The effects of Conscientiousness and Extraversion, however, suggest a new dimension for public policy: encouraging the development of children's Conscientiousness and Extraversion at home and at school would contribute to both health and longevity.

References

- Aitken, M. and D. B. Rubin (1985). Estimation and hypothesis testing in finite mixture models. *Journal of Royal Statistical Society, Series B* 47 (1), pp. 67–75.
- Albouy, V. and L. Lequien (2009). Does compulsory education lower mortality? *Journal of Health Economics* 28, 155–168.
- Allensworth, D., E. Lawson, L. Nicholson, and J. Wyche (Eds.) (1997). *Schools and Health. Our Nation's Investment*. National Academy Press, Washington D.C.
- Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 1, pp. 1–181. Amsterdam: Elsevier B. V.
- Amin, V., J. R. Behrman, and H.-P. Kohler (2015). Schooling has smaller or insignificant effects on adult health in the US than suggested by cross-sectional associations: New estimates using relatively large samples of identical twins. *Social science & medicine* 127, 181–189.
- Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 5, Berkeley, CA, pp. 111–150. University of California Press.
- Apouey, B. and A. E. Clark (2015). Winning big but feeling no better? The effect of lottery prizes on physical and mental health. *Health Economics* 24, 516–538.
- Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review* 24(2), 149–160.

- Arias, E. (2012, September). United States life tables, 2008. *National Vital Statistics Reports* 61(3).
- Auld, M. C. and N. Sidhu (2005, October). Schooling, cognitive ability and health. *Health Economics* 14(10), 1019–1034.
- Batty, G. D., I. J. Deary, and L. S. Gottfredson (2007). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, 278–288.
- Batty, G. D., C. R. Gale, P. Tynelius, I. J. Deary, and F. Rasmussen (2009). IQ in early adulthood, socioeconomic position, and unintentional injury mortality by middle age: a cohort study of more than 1 million Swedish men. *American Journal of Epidemiology* 169(5), 606–615.
- Batty, G. D., K. Modig Wennerstad, G. Davey Smith, D. Gunnell, I. J. Deary, P. Tynelius, and F. Rasmussen (2007). IQ in early adulthood and later cancer risk: cohort study of one million Swedish men. *Annals of Oncology* 18(1), 21–28.
- Batty, G. D., K. M. Wennerstad, G. D. Smith, D. Gunnell, I. J. Deary, P. Tynelius, and F. Rasmussen (2009). IQ in early adulthood and mortality by middle age: cohort study of 1 million Swedish men. *Epidemiology*, 100–109.
- Becker, G. S. (2007, July). Health as human capital: synthesis and extensions. *Oxford Economic Papers* 59(3), 379–410.
- Behrman, J. R., H.-P. Kohler, V. M. Jensen, D. Pedersen, I. Petersen, P. Bingley, and K. Christensen (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography* (48), 1347–1375.
- Bijwaard, G., H. van Kippersluis, and J. Veenman (2015). Education and health: The role of cognitive ability. *Journal of Health Economics* 42 (2015) 29–43, 29–43.
- Bijwaard, G. E. and A. M. Jones (2019). An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality. *Empirical Economics*, 129–175.
- Bijwaard, G. E., M. Myrskylä, P. Tynelius, and F. Rasmussen (2017). Educational gains in cause-specific mortality: Accounting for cognitive ability and family-level confounders using propensity score weighting. *Social Science & Medicine* 184, 49–56.
- Bijwaard, G. E., F. van Poppel, P. Ekamper, and L. Lumey (2015). Gains in life expectancy associated with higher education in men. *PloS one* 10(10), e0141200.
- Bolyard, A. and P. A. Savelyev (2020). Understanding the education polygenic score and its interactions with ses in determining health in young adulthood. SSRN Working Paper <http://dx.doi.org/10.2139/ssrn.3397735>.
- Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources* 43(4), 972–1059.

- Bowen, W. G., M. M. Chingos, and M. S. McPherson (2009). *Finishing College at Public Universities*, pp. 32–56. Princeton University Press.
- Brown, J., J. B. Liebman, and J. Pollet (2002, January). *Appendix. Estimating Life Tables That Reflect Socioeconomic Differences in Mortality*. In: *The Distributional Aspects of Social Security and Social Reform*. University of Chicago Press.
- Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2016). The Effect of College Education on Mortality. *Journal of Health Economics*, 99–114.
- Calvin, C. M., I. J. Deary, C. Fenton, B. A. Roberts, G. Der, N. Leckenby, and G. D. Batty (2011). Intelligence in youth and all-cause-mortality: Systematic review with meta-analysis. *International Journal of Epidemiology* 40(3), 626–644.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Castleman, B. L. and L. C. Page (2014). A trickle or a torrent? Understanding the extent of summer “melt” among college-intending high school graduates. *Social Science Quarterly* 95(1), 202–220.
- Cesarini, D., E. Lindqvist, R. Ostling, and B. Wallace (2016). Wealth, health, and child development: Evidence from administrative data on Swedish lottery players. *The Quarterly Journal of Economics* 131, 687–738.
- Chapman, B. P., B. Roberts, and P. Duberstein (2011). Personality and longevity: Knowns, unknowns, and implications for public health and personalized medicine. *Journal of Aging Research* 2011, 1–24.
- Clark, D. and H. Royer (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review* 103(6), 2087–2120.
- Cobb-Clark, D. A., S. C. Kassenboehmer, and S. Schurer (2014). Healthy habits: the connection between diet, exercise, and locus of control. *Journal of Economic Behavior and Organization* 98, 1–28.
- Conti, G. and J. J. Heckman (2010). Understanding the early origins of the education-health gradient: A framework that can also be applied to analyze gene-environment interactions. *Perspectives on Psychological Science* 5(5), 585–605.
- Conti, G., J. J. Heckman, and S. Urzúa (2010). The education-health gradient. *American Economic Review: Papers and Proceedings* 100(2), 234–238.
- Currie, J. and E. Moretti (2003, November). Mother’s education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics* 118(4), 1495–1532.
- Cutler, D. M., K. Ghosh, and M. B. Landrum (2014). Evidence for significant compression of morbidity in the elderly U.S. population. In D. Wise (Ed.), *Discoveries in the Economics of Aging*, pp. 21–51. University of Chicago Press.

- Cutler, D. M. and A. Lleras-Muney (2008). Education and health: Evaluating theories and evidence. In R. F. Schoeni, J. S. House, G. A. Kaplan, and H. Pollack (Eds.), *Making Americans Healthier: Social and Economic Policy as Health Policy*, pp. 29–60. New York: Russell Sage Foundation.
- Darden, M., D. B. Gilleskie, and K. Strumpf (2018). Smoking and mortality: New evidence from a long panel. *International Economic Review* 59(3), 1571–1619.
- Darden, M. E. and N. W. Papageorge (2018, December). Rational self-medication. NBER Working Paper 25371, National Bureau of Economic Research, Inc.
- de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics* 26(5), 877–895.
- de Walque, D. (2010). Education, information, and smoking decisions: Evidence from smoking histories in the United States, 1940–2000. *Journal of Human Resources* 45(3), 682–717.
- Friedman, H. S. (2000). Long-term relations of personality and health: Dynamism, mechanisms, tropisms. *Journal of Personality* 68(6), 1089–1107.
- Friedman, H. S., M. L. Kern, and C. A. Reynolds (2010). Personality and health, subjective well-being, and longevity. *Journal of Personality* 78(1), 179–216.
- Friedman, H. S., J. S. Tucker, J. E. Schwartz, L. R. Martin, C. Tomlinson-Keasey, D. L. Wingard, and M. H. Criqui (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology* 68(4), 696–703.
- Friedman, H. S., J. S. Tucker, C. Tomlinson-Keasey, J. E. Schwartz, D. L. Wingard, and M. H. Criqui (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology* 65(1), 176–185.
- Frijters, P., J. P. Haisken-DeNewb, and M. A. Shields (2005). The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics* 24, 997–1017.
- Galama, T. and H. van Kippersluis (2015). A theory of education and health. CESR-Schaeffer Working Paper No 2015-001.
- Galama, T. J., A. Lleras-Muney, and H. van Kippersluis (2018, September). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. *The Oxford Research Encyclopedia, Economics and Finance* (oxfordre.com/economics), 1–96.
- Gardner, J. and A. J. Oswald (2007). Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of Health Economics* 26, 49–60.
- Gensowski, M. (2018). Personality, IQ, and lifetime earnings. *Labour Economics* 51, 170–183.
- Gilleskie, D. (2014). Dynamic models: Econometric considerations of time. In *Encyclopedia of Health Economics, Volume 1*, pp. 209–216. Elsevier Inc.

- Grimard, F. and D. Parent (2007, September). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics* 26(5), 896–926.
- Grossman, M. (2000). The human capital model. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1, Chapter 7, pp. 347–408. Amsterdam: Elsevier Science B. V.
- Grossman, M. (2015). The relationship between health and schooling: What’s new? *Nordic Journal of Health Economics* 3(1), 7–17.
- Harris, K. and S. Vazire (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass* 10(11), 647–667.
- Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health and smoking. *Journal of Political Economy* 126, 197246.
- Heckman, J. J. and T. Kautz (2014). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, Chapter 9, pp. 341–430. Chicago: University of Chicago Press.
- Heckman, J. J. and P. A. LaFontaine (2010, May). The American high school graduation rate: Trends and levels. *Review of Economics and Statistics* 92(2), 244–262.
- Heckman, J. J. and B. S. Singer (1984, March). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2), 271–320.
- Heckman, J. J., J. Stixrud, and S. Urzúa (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24(3), 411–482.
- Hong, K., P. Savelyev, and K. Tan (2020). Understanding the mechanisms linking education with longevity. *Journal of Human Capital* (forthcoming).
- Hsieh, C.-S. and H. van Kippersluis (2018). Smoking initiation: Peers and personality. *Quantitative Economics* 9, 825–863.
- Huang, G.-H. and K. Bandeen-Roche (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika* 69(1), 5–32.
- John, O. P. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2 ed.), Chapter 4, pp. 102–138. New York: The Guilford Press.
- Kenkel, D. S. (2000). Prevention. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (1 ed.), Volume 1, Chapter 31, pp. 1675–1720. Elsevier.

- Kern, M. L. and H. S. Friedman (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology* 27(5), 505–512.
- Kim, B. and C. J. Ruhm (2012). Inheritances, health and death. *Health Economics* 21, 127–144.
- Kippersluis, H. v. and T. J. Galama (2014). Wealth and health behavior: Testing the concept of a health cost. *European Economic Review* 72, 197–220.
- Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources* 40(1), 144–168.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies* 72(1), 189–221.
- Lochner, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 2, pp. 183–282. Amsterdam: Elsevier B. V.
- Lundborg, P., C. H. Lyttkens, and P. Nystedt (2016). The effect of schooling on mortality: New evidence from 50,000 Swedish twins. *Demography* (53), 1135–1168.
- Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- Madsen, M., A.-M. N. Andersen, K. Christensen, P. K. Andersen, and M. Osler (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology* 172.
- Martin, L. R. and H. S. Friedman (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality* 68(1), 85–110.
- Martin, L. R., H. S. Friedman, and J. E. Schwartz (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology* 26 (4), 428–436.
- Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives* 32(2), 2–16.
- Meghir, C., M. Palme, and E. Simeonova (2018, April). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics* 10(2), 234–56.
- Mroz, T. A. (1999). Discrete factor approximation in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics* 92(2), 233–274.
- Murphy, K. M. and R. H. Topel (2006, October). The value of health and longevity. *Journal of Political Economy* 114(5), 871–904.

- Murray, D. W., S. H. Kollins, K. K. Hardy, H. B. Abikoff, J. M. Swanson, C. Cunningham, B. Vitiello, M. A. Riddle, M. Davies, L. L. Greenhill, J. T. McCracken, M. J. J., P. K., A. Skrobala, W. T., W. S. B., G. J. K., and S. Z. Chuang (2007). Parent versus teacher ratings of attention-deficit/hyperactivity disorder symptoms in the preschoolers with attention-deficit/hyperactivity disorder treatment study (PATS). *J. Child Adolesc. Psychopharmacol.* 17 (5), 605–20.
- Næss, Ø., D. A. Hoff, D. Lawlor, and L. H. Mortensen (2012). Education and adult cause-specific mortality—examining the impact of family factors shared by 871 367 Norwegian siblings. *International Journal of Epidemiology* 41(6), 1683–1691.
- Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg (2007, December). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science* 2(4), 313–345.
- Ross, C. E. and C.-I. Wu (1995). The links between education and health. *American Sociological Review* 60(5), 719–745.
- Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.
- Savelyev, P. A. and K. Tan (2019). Socioemotional skills, education and health-related outcomes of high-ability individuals. *American Journal of Health Economics* 5(1), 250–280.
- Savelyev, P. A., B. Ward, R. Krueger, and M. McGue (2020). Health endowments, schooling allocation in the family, and longevity: Evidence from US twins. HCEO and SSRN Working Paper <http://dx.doi.org/10.2139/ssrn.3193396>.
- Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- Schwandt, H. (2018). Wealth shocks and health outcomes: Evidence from stock market fluctuations. *American Economic Journal: Applied Economics* 10(4), 349–77.
- Sears, R. R. (1984). The Terman gifted children study. In S. Mednick, M. Harway, and K. Finello (Eds.), *Handbook of Longitudinal Research*, pp. 398–414. New York: Praeger.
- Snyder, S. E. and W. N. Evans (2006). The effect of income on mortality: Evidence from the social security notch. *The Review of Economics and Statistics* 88, 482–495.
- Terman, L. M. (1986). Terman life-cycle study of children with high ability by Terman L. M. et al., 1922-1986 [computer file]. 2nd ICPSR release. Palo Alto, CA: Robert R. Sears [producer], 1986. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1989. doi:10.3886/ICPSR08092.
- Terman, L. M., B. T. Baldwin, E. Bronson, J. C. DeVoss, F. Fuller, T. Lee Kelley, M. Lima, H. Marshall, A. H. Moore, A. S. Raubheimer, G. M. Ruch, R. L. Willoughby, J. Benson Wyman, and D. Hazeltine Yates (1925). *Genetic Studies of Genius: Mental and Physical Traits of a Thousand Gifted Children*, Volume 1. Stanford University, CA: Stanford University Press.

- Terry, L. and co-authors (1964). Smoking and health: Report of the advisory committee to the surgeon general of the United States. Public Health Service publication No. 1103. Technical report, Department of Health, Education, and Welfare.
- Van Den Berg, G., L. Janys, and K. Christensen (2015). The effect of education on mortality. IZA working paper.
- Van den Berg, G. J. (2001). Duration models: Specification, identification and multiple durations. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbooks in Economics*, pp. 3381–3460. New York: North-Holland.
- Van Der Pol, M. (2011). Health, education, and time preference. *Health Economics* 20, 917–929.
- van Kippersluis, H., O. O'Donnell, and E. van Doorslaer (2011). Long-run returns to education: Does schooling lead to an extended old age? *The Journal of Human Resources* 46, 695–721.
- Vollrath, M., M. Landolt, H. Gnehm, J. Laimbacher, and F. H. Sennhauser (2007). Child and parental personality are associated with glycaemic control in type 1 diabetes. *Diabetic Medicine* 24(9), 1028–1033.

[FOR ONLINE PUBLICATION]

Web Appendix to “Conscientiousness,
Extraversion, College Education, and Longevity
of High-Ability Individuals”¹

Peter A. Savelyev

The College of William & Mary and IZA

June 10, 2020

¹A version of this paper published by the *Journal of Human Resources* is available from <https://doi.org/10.3368/jhr.58.1.0918-9720R2>. This paper was presented to The Health Economics Workshop at the NBER Summer Institute; The Annual Meeting of the American Economic Association in Chicago; The European Economic Association and Econometric Society Annual European Meeting in Gothenburg, Sweden; The Health Economics Workshop and the Labor Group Seminar at the University of Chicago; Applied Microeconomics seminar at the University of North Carolina; The Institute on Health Economics, Health Behaviors, and Disparities at Cornell University; Economics seminar of the Andrew Young School of Policy Studies, Georgia State University; Empirical Micro seminar of the University of Houston and Rice University; RAND Labor and Population Seminar in Santa-Monica; Departmental Seminar in Economics at Vanderbilt University; the Applied Microeconomics Seminar at Vanderbilt Law School; The Quantitative Methods Brown Bag at the Department of Psychology and Human Development, Peabody College, Vanderbilt University; David Eccles School of Business research seminar, University of Utah; and the Upjohn Institute for Employment Research seminar. I thank participants of these meetings for useful suggestions and stimulating discussions. I am grateful to Gary Becker, Gabriella Conti, Miriam Gensowski, Mike Grossman, Tim Kautz, Don Kenkel, Adriana Lleras-Muney, Willard Manning, David Meltzer, Rémi Piatek, Ben Williams, and especially Jim Heckman for productive comments. This version of the paper directly benefited from research assistance provided by Atticus Bolyard and Max Sacher and from proofreading by Xiaoyu (Nancy) Chen, Renee Garrow, Isabel Harber, and Eli Rothleder. Work on this paper stretched over a decade, over which many RAs contributed to its development, namely Mihir Gandhi, Kai Hong, Hanchen Jiang, Adam Shriver, Ivana Stosic, and Kegon Tan. Summer interns Keith Dent, Son Nghiem, John Spraul, and Cody Vaughn used previous versions of this paper to learn empirical skills and provided help. The Terman data are provided by the Interuniversity Consortium for Political and Social Research, Ann Arbor, MI. Peter Savelyev gratefully acknowledges research support from the College of William & Mary, Vanderbilt, and the ERC at the University of Chicago. An early version of this research was

Contents

A Exploratory and Confirmatory Factor Analysis of Noncognitive Measures used in the Terman Study	2
A.1 Establishing a Set of Childhood Noncognitive Measures that are Conceptually Related to the Big Five	4
A.2 Using Multiple Methods to Establish the Number of Latent Factors that Explain the Given Set of Measures	4
A.3 Using EFA to Finalize the Number of Factors and Identify Clusters of Measures that Proxy these Factors	5
A.4 Using CFA to Confirm the EFA Results	6
A.5 Using the 1940 Data to Augment the 1922 Factors	11
B Supplementary Figures and Tables	20
C The Likelihood Function	29
References	30

supported by the Merck Quantitative Science Graduate Fellowship in Health Economics. The views expressed in this paper are those of the author and may not coincide with those of the funders.

A Exploratory and Confirmatory Factor Analysis of Noncognitive Measures used in the Terman Study

The aim of this section is to reanalyze noncognitive measures from the Terman data in order to obtain five latent factors that are related to the contemporary Big Five traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

According to [John and Srivastava \(1999\)](#), Openness describes the breadth, depth, originality, and complexity of an individual’s mental and experiential life; Conscientiousness represents “individual differences in the propensity to follow socially prescribed norms for impulse control, to be task- and goal- directed, to be planful, to delay gratification, and to follow norms and rules.” Alternatively, Conscientiousness can be described as the “propensity to be organized, controlled, industrious, responsible, and conventional” ([Roberts et al., 2009](#)). Extraversion implies an energetic approach to the social and material world and includes facets such as sociability, activity, assertiveness, and positive emotionality. Agreeableness entails a prosocial and communal orientation towards others and includes facets such as altruism, tender-mindedness, trust, and modesty. Finally, Neuroticism contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, or tense.

[Martin and Friedman \(2000\)](#) show that strong links exist between personality measures used in the Terman Data and the Big Five factors. However, previous authors documented neither their exploratory nor their confirmatory factor analyses, which motivates research documented in this appendix: finding the best specification for the latent factor model to be used in the main paper.

As a result of the analysis presented below, I find that four of five factors defined in this paper, childhood Conscientiousness and Extraversion as well as adult Agreeableness and Neuroticism, are defined in a similar way to factors that were empirically shown to be correlated with the Big Five factors having the same names ([Martin and Friedman, 2000](#)). In particular, Martin and Friedman show that Conscientiousness correlates well with Conscientiousness from the Big Five taxonomy (correlation 0.55, and p -value below 0.001). The authors also demon-

strate a relationship between their Terman Sociability factor and Extraversion from the Big Five taxonomy (with correlation 0.40, and p -value below 0.001). Extraversion in this paper is defined similarly to Martin and Friedman's Sociability, and, hence, should also be related to Big Five Extraversion. Openness, as defined in this paper, was not empirically linked to the Big Five, but it can be linked to the Big Five theoretically. To achieve the best model fit I allow for cross-loadings (to be defined below), instead of using dedicated measures, which were implemented in previous research. However, the effects of using cross-loadings on final model results are minor.

The plan of the analysis described in this section is as follows:

1. Select a set of high-quality childhood personality measures from the Terman data so that each measure from the set is conceptually related to the Big Five taxonomy.
2. Use multiple methods to establish the number of latent common factors that explain the set of measures identified in (1). (These methods usually give several different answers to be considered in the further analysis).
3. For each number of factors suggested by step (2), perform the exploratory factor analysis (EFA) in order to establish the most reasonable number of factors; conditional on the most reasonable number of latent factors, find clusters of measures that proxy the factors.
4. Use confirmatory factor analysis (CFA) to test model specifications suggested by EFA. (As shown below, I establish measures that define childhood Conscientiousness, Openness, and Extraversion).
5. Use the 1940 data to proxy those Big Five factors that are not represented in the 1922 data. (As described below, I add adulthood Agreeableness and Neuroticism.)

Each of the five steps in the above outline corresponds to one of the five sections below.

A.1 Establishing a Set of Childhood Noncognitive Measures that are Conceptually Related to the Big Five

We want to find clusters of childhood noncognitive measures from 1922 that proxy Big Five factors. Childhood measures are more valuable than adulthood ones for the purpose of this paper, as they are measured before subjects start college. Additionally, the 1922 measures are based on ratings from teachers and parents, which makes them more reliable than self-ratings from 1940.

Based on theoretical definitions of the Big Five provided above, results from previous research (Martin and Friedman, 2000; Martin et al., 2007), and advice from a leading psychologist¹, I use the following childhood measures that are potentially related to the Big Five: prudence and forethought, conscientiousness, truthfulness, desire to know, originality, general intelligence, fondness for a large group, leadership, popularity with other children, cheerfulness and optimism, permanency of moods, sensitiveness to approval and disapproval, freedom from vanity and egoism, sympathy and tenderness, generosity and unselfishness.² The next section proceeds with establishing the number of latent common factors contained in these correlated measures.

A.2 Using Multiple Methods to Establish the Number of Latent Factors that Explain the Given Set of Measures

I use five alternative methods that establish the number of latent factors behind the measures identified above. I use (1) the Guttman-Kaiser rule, (2) the scree test (e.g., Gorsuch, 1983), (3) the Horn method (Gorsuch, 1983; Guttman, 1954; Horn, 1965; Kaiser, 1960, 1961), (4) the Minap method (Velicer, 1976), and (5) the Onatski (2009) procedure. The “optimal number of factors” represents an optimal balance between minimizing the number of common factors

¹I am grateful to Angela Duckworth, a psychologist from the University of Pennsylvania, for a fruitful discussion about conceptual links between Terman measures and the Big Five.

²See Figure A-1 for an example of a question measuring prudence. Teachers and parents had to put a cross anywhere on the line.

and maximizing the explained variance in the data. Viewing each measure as a separate factor would explain 100% of variance but would also imply no dimensionality reduction and would lead to problems such as high measurement error and multicollinearity. Instead, I want to establish a small set of latent factors that explain a substantial share of variance, such that adding an additional factor to this optimal set would have a negligible impact on explaining the variance of measures.

Methods mentioned above differ by their approach to solving this optimization task, leading to some variation in the recommended optimal number of factors (see also Heckman, Pinto, and Savelyev (2013)). As Table A-1 shows, the optimal number of factors is estimated to be 2, 3, and 4.³ With the help of the EFA analysis in the next section I argue that 3 is the most reasonable choice.

A.3 Using EFA to Finalize the Number of Factors and Identify Clusters of Measures that Proxy these Factors

EFA can only be performed conditional on a given number of factors. When there are several options (in our case 2, 3, or 4), it is standard to perform EFA conditional on each number, compare results, weigh pros and cons, and finalize the number of factors.⁴

I use the EFA based on the Quartimin rotation, an established and widely used oblique EFA procedure with a clear mathematical interpretation that we discuss in Web Appendix H to Heckman, Pinto, and Savelyev (2013). In the same paper we argue that the choice of measures for factors is robust to the choice of alternatives to the Quartimin.

I begin by performing the EFA for all 15 measures listed above. Results for males and females are presented in Table A-2. Strong loadings, those that are both statistically significant at a 5% level and are above 0.6, are in bold font.⁵ Conditional on two factors we can see two clear

³See also the corresponding scree plots in Figure A-2.

⁴Following Friedman et al. (1995, 1993), I use averaged teachers' and parents' ratings.

⁵Heckman, Pinto, and Savelyev (2013) used 0.6 as a threshold for separating strong loadings from weak loadings. The threshold is traditional and useful for exploratory purposes, but still arbitrary, so it should be interpreted as such.

clusters: cluster 1 is mostly related to measures that resemble Conscientiousness (prudence and forethought, conscientiousness, truthfulness), and cluster 2 is mostly related to measures that resemble Extraversion (fondness for large groups, leadership, popularity with other children). Adding a third factor preserves the above-mentioned clusters, but adds another cluster that resembles Openness (desire to know, originality, and intelligence). While adding a third factor clearly improves the model, there are no clear benefits associated with adding a fourth factor. For men, sympathy and generosity resemble a cluster, but this finding is not true for women.

The next step of EFA is to drop measures that do not clearly belong to any cluster and to analyze the resulting loadings again to check whether patterns of measures are preserved after less relevant measures are excluded from the analysis. I drop all such less relevant measures but two: “freedom from vanity” and “sympathy.” I keep these measures since they show some association with the Conscientiousness cluster and because “freedom from vanity” was used by Friedman and co-authors as a measure of Conscientiousness. The table for the second stage, [A-3](#), shows that results of the first stage are preserved and that “freedom from vanity” and “sympathy” are mostly related to the Conscientiousness cluster, but their link to Conscientiousness tends to be weaker than that of other measures. While keeping “freedom from vanity” and “sympathy” in mind as candidates for the Conscientiousness cluster with the caveat that they are both less correlated and less theoretically linked, I also perform the third stage without these two measures for the case of three factors leaving checks for the two questionable measures to the CFA stage. The third-stage CFA for a three-factor model shown in [Table A-4](#) again confirms patterns found in the first stage. The table also shows good fit statistics, which I discuss below since fit statistics are also relevant for the final decision based on CFA.

A.4 Using CFA to Confirm the EFA Results

Unlike EFA, which places no restrictions on loadings, CFA requires a number of exclusion restrictions necessary for the factor model identification. CFA is based on a factor model, which also serves as the measurement system for the main estimated model of the paper upon

being confirmed as acceptable.

One straightforward exclusion restriction would be to set all non-diagonal loadings to zero (the diagonal loadings are in bold in Table A-4). This approach is appealing for this data set since the non-diagonal elements are all substantially smaller than the diagonal elements. Hence, setting non-diagonal elements to zero should not introduce any sizable model misspecification. This approach leads to so-called “dedicated measures,” such that each measure proxies one and only one factor. However, some non-diagonal elements in Table A-4 are still statistically significant and reach up to 0.31 by absolute value. The assumption behind our identifying exclusion restriction would be less restrictive if non-diagonal elements that are distinguishable from zero were allowed to differ from zero as well.

There are two potential caveats. First, it is not always possible to both achieve identification and allow for all sizable non-diagonal elements to be unrestricted. Fortunately, for this particular data, this less restrictive approach leads to an identifiable model.

Second, cross-loadings may undermine the interpretation of factors (see, e.g., Kline, 2011), but when cross-loadings are small and sensible, as in this paper, this should not be an issue. Indeed, cross-loadings in my analysis are much smaller than diagonal loadings. For instance, in Table A-5, model A for males, diagonal loadings take values from 0.563 to 0.894 and are precisely determined, while cross-loadings take values from -0.086 to 0.214 and are much less precisely determined. Thus, the interpretation of factors is still derived from the diagonal loadings just as in a model with dedicated measures, while cross-loadings can be interpreted as naturally-occurring minor correlations between certain common factors and measures of certain other common factors.

Moreover, the cross-loadings can be reasonably interpreted given that the interpretation of the common factor is derived from the diagonal loadings. Indeed, one of the largest cross-loadings in Panel A for males is 0.186 for “popular with other children” as a weak proxy of Conscientiousness. We can argue that there is no contradiction that someone who is at the same time more prudent, more conscientious, and more truthful (which are the main measures of Conscientiousness defined in this paper) would also gain additional popularity with other

children. Similarly, there is no contradiction that desire to know also somewhat correlates with Conscientiousness, as we expect conscientious people to look ahead and thus appreciate the value of education (a small cross-loading of 0.065). Cross-loadings for Openness range from -0.086 to 0.214 and include positive loadings for prudence and leadership and a negative one for popularity. Interpretations of these small correlations are somewhat less straightforward, but it is reasonable that a more original and intelligent person would also be better at planning and leadership, while at the same time they may lose some popularity for being too different from others (originality). Finally, extraversion has a small positive correlation with truthfulness (loading 0.055) and a small negative correlation with the desire to know (-0.086). The negative correlation is clear: extraverted people like socializing, which may limit spending time studying, leading to smaller rating on the “desire to know,” since raters may associate the desire to know with the amount of studying. Truthfulness might be a trait that is valued by people one socializes with, possibly leading to a positive correlation with extraversion. The sign of cross-loadings is preserved for females. Thus, cross-loadings that I allowed for milder exclusion restrictions are small and do not contradict the interpretation of factors that is based on the diagonal loadings.

I compare panel A in Table A-5 representing a model with cross-loadings based on stage three of the EFA with three other panels: panel B represents a model with dedicated measures only; panel C is a model with “freedom from vanity,” a questionable measure, allowed to affect any factor as the least restrictive case to consider; and panel D, with “sympathy,” another questionable measure, allowed to affect any factor. As I argue above, the fit measures are good or fair for models B, C, and D, but model A shows the best fit and therefore I choose it as my main model.

Goodness of Fit Goodness of fit statistics are described in many sources, such as [Bollen and Long \(1993\)](#), [Kline \(2011\)](#), and [Kolenikov \(2009\)](#). Here I provide a short intuitive summary and discuss the interpretation of results.

The Likelihood Ratio Test The likelihood ratio (LR) test, which is often referred to in psychometric literature as “the Chi-square test,” compares two models: the constrained model, which is the factor model that we test, and the unconstrained (unstructured, or saturated) model, in which the model can exactly predict each observed covariance. Ideally, restrictions of the factor model do not matter, so the test will not be rejected. In practice, this test is usually rejected, leading some authors to doubt the reliability of the test (e.g., [Bentler and Bonett, 1980](#); [Du and Tanaka, 1989](#)) and ignore results of the test as uninformative or unreliable. Current literature increasingly recognizes that the test should be taken seriously (see, e.g., [Kline, 2011](#), for a survey).

While I agree with the more recent claims that the test is generally reliable and informative for the question being tested, I view frequent rejection of the test, even for otherwise clean models, as a reasonable price for the major benefits of dimensionality reduction. We know that statistics is an art of reasonable approximation, not exact fit, in line with which dimensionality reduction techniques come at a cost of multiple model restrictions.⁶ The expectation that the test will not be rejected is the expectation that a model with greatly reduced dimensionality will still be exactly as informative as a fully saturated model, allowing for all possible correlations between all possible measures. It would be over-optimistic to expect that none of these restrictions will lead to at least some simplifications of reality.

My point is much in line with [Steiger \(2007\)](#) and [Miles and Shevlin \(2007\)](#). In particular, Steiger notes that the chi-square test is not only about testing a hypothesis of a perfect fit, which we know is always false, but also has the logical weakness inherent to the accept-support tests: researchers who use tests with low power (say, those conducting studies with low sample size) tend to accept the hypothesis and so support their model, while those using high power tests tend to reject it and therefore have trouble defending their analysis. These properties make the LR test not particularly useful, and so I do not use the LR test in this paper.

⁶A relevant quotation from statistician George Box is “Essentially, all models are wrong, but some are useful” or another version, “...all models are wrong; the practical question is how wrong do they have to be to not be useful” (Box, G. E. P., and Draper, N. R., (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York, NY).

The other fit statistics are called “approximate indexes.” Unlike the LR test, these are not true statistical tests, but they are informative of the fit (Miles and Shevlin, 2007). Thresholds (or “rules of thumb”) for these indices are recommended based on Monte-Carlo studies (Kline, 2011). I give a short overview of indices that I use below. For more detailed information and formulas see, e.g., Bollen and Long (1993), Kline (2011), and Kolenikov (2009).

SRMR The Standardized Root-Mean-Square Residual (SRMR) is an absolute measure of fit. It is a mean square root of a squared standardized difference between a true element of the variance-covariance matrix and an element of the same matrix as predicted by the model. Clearly, the smaller the SRMR, the better the model fit. The rule of thumb for a good fit is $SRMR \leq 0.08$, which is satisfied for all models in Table A-5.

RMSEA Another index is the Root Mean Square Error of Approximation (RMSEA), which is classified as a parsimony index. RMSEA is proportional to the square root of the likelihood ratio from the LR test discussed above, so a lower number implies a better fit. RMSEA is also inversely proportional to the square root of the number of model restrictions. If the number of model restrictions is small, then RMSEA may become larger because of a decreased denominator, which is a punishment for a non-parsimonious model. The rule of thumb is $RMSEA \leq 0.05$, which is true for all models of types A and D in Table A-5.

CFI and TLI Finally, I use two comparative indices: the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). These indices are relative because they are based on a comparison of two likelihood ratios: (1) Target (constrained) model relative to unconstrained (saturated) model; (2) Independence model (diagonal variance-covariance matrix) relative to the target (constrained) model. Since measures are usually highly correlated, we expect our target factor model to have much better fit than the independence model, which makes the likelihood ratio (2) much larger than (1) and drives both indices close to 1. The rule of thumb is that CFI and $TLI > 0.9$. In Table A-5 this is the case for all models of type A, C, and D.

To conclude, comparison of all approximate fit indices described above suggests that Model

A is preferable to model B, and that model A satisfies all recommended thresholds. Also, there is no consistent evidence that models C and D are better than model A. Fit statistics for models C and D tend to be worse or comparable to those for model A. Given doubts about theoretical fit of “freedom from vanity” and “sympathy” to Conscientiousness, I chose model A as the most reasonable model.

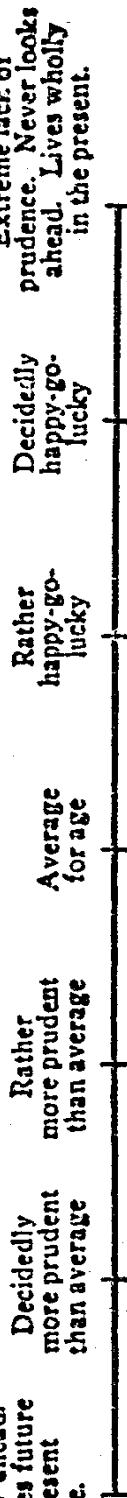
A.5 Using the 1940 Data to Augment the 1922 Factors

Analysis using the 1940 items that allow an inclusion of additional Big Five factors is documented in the Web Appendix to a companion paper by [Savelyev and Tan \(2019\)](#). Table [A-6](#) summarizes results of the EFA and CFA of the companion paper. As mentioned in the main text, I do not use the five-factor model as my preferred specification.

Figure A-1: A Photocopy of One Question from the Original 1922 Questionnaire: Measuring Prudence and Forethought

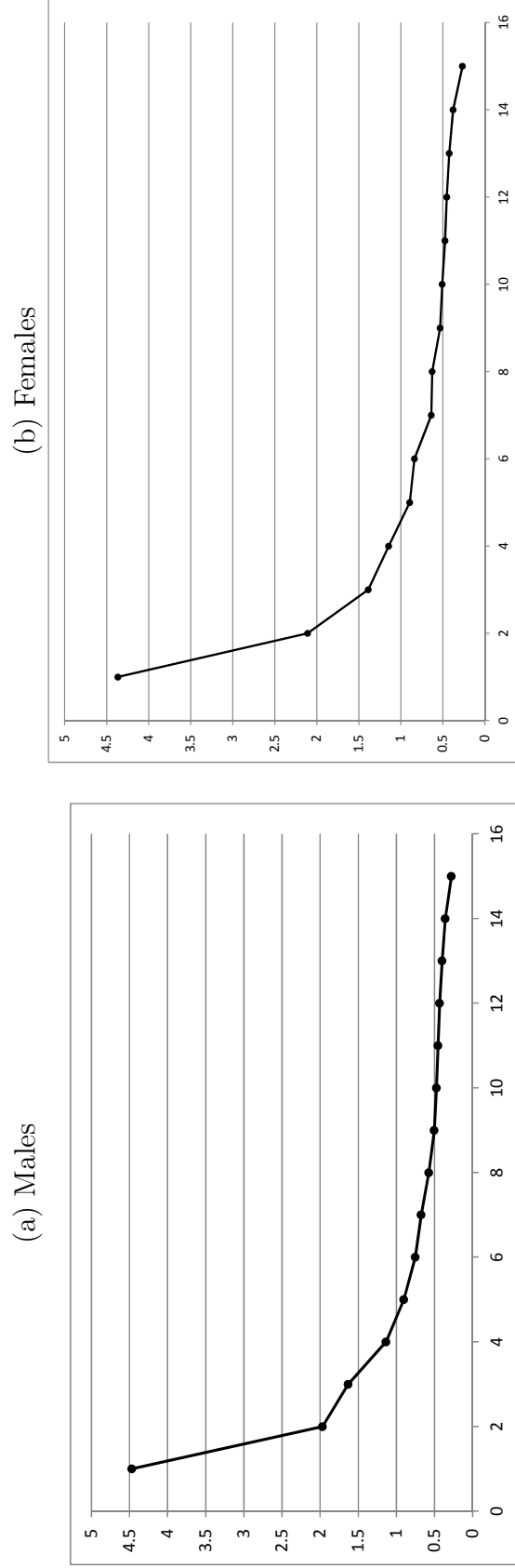
Trait 3. Prudence and forethought.

Extraordinary prudence.
Always looks ahead.
Never sacrifices future
good for present
pleasure.



Source: [Terman and Sears \(2002\)](#). Respondents were asked to put a mark anywhere on this line.

Figure A-2: Scree Plots



Notes: Scree plots are shown for 15 items potentially related to the Big Five traits. Calculations are based on the Terman data.

Table A-1: Estimates of the Number of Factors Based on Various Methods

Procedure	Males	Females
Scree ^(a)	2	3
Guttman-Kaiser ^(b)	4	4
Horn ^(c)	4	3
Minap ^(d)	3	2
Onatski ^(e)		
3-7 factors	3	-
2-7 factors	2	-
1-7 factors	2	-

Notes:

^(a)Scree test is by [Cattell \(1966\)](#). The corresponding scree plots are documented in Figure [A-2](#).

^(b)See [Gorsuch \(1983\)](#); [Guttman \(1954\)](#); [Kaiser \(1960, 1961\)](#) for a discussion of the Guttman-Keiser rule. The method is known to exclude factors that clearly have little explanatory power, but it often overestimates the number of informative factors ([Zwick and Velicer, 1986](#)).

^(c)Horn's (1965) parallel analysis procedure.

^(d)See [Velicer \(1976\)](#) for the Minap description.

^(e)We apply Onatski's (2009) procedure at the 5% level of significance for a minimum of three, two, and one factors and for a maximum of seven factors. [Onatski \(2009\)](#) warns that the asymptotic approximation may be poor in a case where sample size is small and the number of measures is low. "-" denotes that [Onatski's](#) algorithm does not converge to any number in the given range.

Table A-2: First Stage of the EFA

	2 factors		3 factors			4 factors			
	1	2	1	2	3	1	2	3	4
Males									
Prudence and forethought	0.697*	-0.086*	0.623*	0.160*	-0.119*	0.722*	0.112*	0.038	-0.143*
Conscientiousness	0.862*	-0.054*	0.827*	0.076*	-0.066*	0.775*	0.060*	-0.015	0.112*
Truthfulness	0.771*	0.012	0.768*	0.031	0.008	0.673*	0.031	0.015	0.170*
Desire to know	0.371*	0.143*	0.076*	0.721*	-0.041	0.075*	0.719*	-0.070*	0.034
Originality	0.295*	0.233*	0.011	0.679*	0.069*	0.037	0.673*	0.064	-0.019
Intelligence	0.300*	0.241*	0.022	0.680*	0.076*	0.009	0.687*	0.027	0.052
Fondness for large groups	-0.163*	0.730*	-0.147*	0.026	0.717*	-0.261*	0.060	0.566*	0.227*
Leadership	0.039	0.579*	-0.032	0.200*	0.537*	0.055	0.170*	0.653*	-0.156*
Popularity w other children	0.061*	0.699*	0.113*	-0.062*	0.744*	0.098*	-0.094*	0.813*	0.031
Optimism	0.064	0.596*	0.006	0.190*	0.538*	-0.137*	0.229*	0.375*	0.277*
Permanency of moods	0.330*	0.228*	0.314*	0.065	0.213*	0.317*	0.049	0.256*	0.004
Sensitiveness	0.270*	0.150*	0.251*	0.066	0.132*	0.118*	0.095*	0.013	0.262*
Freedom from vanity	0.457*	0.050	0.585*	-0.245*	0.125*	0.468*	-0.244*	0.108*	0.217*
Sympathy	0.443*	0.239*	0.453*	0.021	0.231*	0.142*	0.087*	-0.049	0.626*
Generosity	0.404*	0.381*	0.440*	-0.021	0.390*	0.134*	0.035	0.135*	0.606*
Sample size	695		695			695			
Females									
Prudence and forethought	0.587*	0.015	0.562*	0.068	-0.007	0.587*	0.057	0.026	-0.039
Conscientiousness	0.872*	-0.053*	0.845*	0.041	-0.048*	0.868*	0.021	-0.017	0.007
Truthfulness	0.772*	-0.003	0.734*	0.103*	-0.048	0.745*	0.091*	-0.016	-0.005
Desire to know	0.245*	0.332*	0.043	0.680*	-0.054	0.046	0.678*	-0.048	0.002
Originality	0.214*	0.414*	0.045	0.597*	0.075	0.026	0.592*	0.071	0.051
Intelligence	0.258*	0.372*	0.062	0.683*	-0.005	0.049	0.686*	-0.008	0.027
Fondness for large groups	-0.179*	0.639*	-0.133*	-0.020	0.712*	-0.130*	-0.038	0.720*	0.004
Leadership	-0.082*	0.665*	-0.148*	0.347*	0.474*	-0.130*	0.332*	0.483*	-0.019
Popularity w other children	0.129*	0.606*	0.190*	-0.046	0.720*	0.168*	-0.054	0.718*	0.035
Optimism	0.067	0.572*	0.043	0.215*	0.460*	0.052	0.205*	0.479*	-0.023
Permanency of moods	0.288*	0.200*	0.300*	0.021	0.200*	0.352*	0.013	0.233*	-0.056
Sensitiveness	0.291*	0.090	0.309*	-0.032	0.136*	0.199*	-0.024	0.089	0.212*
Freedom from vanity	0.572*	-0.104*	0.629*	-0.168*	0.015	0.583*	-0.159*	0.018	0.058
Sympathy	0.474*	0.197*	0.485*	0.028	0.204*	-0.006	0.005	-0.004	1.191*
Generosity	0.517*	0.229*	0.528*	0.042	0.223*	0.373*	0.053	0.167*	0.291*
Sample size	536		536			536			

Notes: An asterisk denotes a 5% level statistical significance.

Table A-3: Second Stage of the EFA

	2 factors		3 factors			4 factors			
	1	2	1	2	3	1	2	3	4
Male									
Prudence and forethought	0.690*	-0.044	0.627*	0.130*	-0.056	0.634*	0.085*	-0.023	-0.265*
Conscientiousness	0.858*	-0.031	0.837*	0.054*	-0.033	0.822*	0.060*	-0.016	-0.033
Truthfulness	0.774*	0.004	0.783*	0.010	0.017	0.781*	0.042	0.004	0.110*
Desire to know	0.387*	0.137*	0.082*	0.710*	-0.040	0.069*	0.740*	-0.061*	0.054
Originality	0.309*	0.223*	0.005	0.690*	0.051	-0.001	0.668*	0.069*	-0.101*
Intelligence	0.319*	0.225*	0.024	0.690*	0.049	0.016	0.700*	0.044	0.045
Fondness for large groups	-0.133*	0.688*	-0.149*	0.073	0.627*	-0.134*	0.079*	0.639*	0.223*
Leadership	0.054	0.643*	-0.042	0.225*	0.567*	-0.030	0.139*	0.657*	-0.254*
Popularity w other children	0.065*	0.737*	0.079*	-0.055*	0.862*	0.118*	-0.077*	0.800*	0.022
Freedom from vanity	0.435*	0.049	0.549*	-0.232*	0.153*	0.560*	-0.219*	0.128*	0.054
Sympathy	0.436*	0.181*	0.422*	0.055	0.162*	0.430*	0.103*	0.135*	0.259*
Sample size	695		695			695			
Female									
Prudence and forethought	0.577*	0.055	0.560*	0.063	0.028	0.461*	0.102	0.029	0.085
Conscientiousness	0.893*	-0.021	0.890*	0.013	0.000	1.029*	-0.025	0.002	-0.028*
Truthfulness	0.770*	0.023	0.738*	0.090*	-0.02	0.514*	0.164*	-0.016	0.219*
Desire to know	0.235*	0.372*	0.041	0.676*	-0.067*	0.059	0.672*	-0.063	-0.062
Originality	0.199*	0.498*	0.041	0.601*	0.094	-0.003	0.617*	0.093*	0.014
Intelligence	0.246*	0.417*	0.053	0.690*	-0.013	-0.017	0.717*	-0.014	0.059
Fondness for large groups	-0.154*	0.548*	-0.119*	-0.025	0.692*	-0.019	-0.064	0.694*	-0.121*
Leadership	-0.091*	0.750*	-0.134*	0.336*	0.516*	-0.018	0.299*	0.522*	-0.174*
Popularity w other children	0.137*	0.532*	0.185*	-0.046	0.729*	0.040	-0.013	0.741*	0.173*
Freedom from vanity	0.552*	-0.129*	0.585*	-0.141*	0.010	0.044	0.000	0.005	0.719*
Sympathy	0.430*	0.179*	0.415*	0.073	0.181*	0.157*	0.155*	0.178*	0.302*
Sample size	536		536			536			

Notes: An asterisk denotes a 5% level statistical significance.

Table A-4: Third Stage of the EFA, a Three-Factor Case

	Males				Females			
	Conscientiousness	Openness	Extraversion	Standard Error	Conscientiousness	Openness	Extraversion	Standard Error
Conscientiousness								
Prudence	0.628 (0.035)	0.108 (0.040)	-0.037 (0.032)	(0.041)	0.541 (0.041)	0.070 (0.051)	0.030 (0.046)	(0.046)
Conscientiousness	0.904 (0.027)	-0.016 (0.018)	-0.002 (0.015)	(0.031)	0.977 (0.031)	-0.028 (0.012)	0.007 (0.013)	(0.013)
Truthfulness	0.740 (0.032)	0.003 (0.031)	0.045 (0.026)	(0.038)	0.682 (0.038)	0.111 (0.042)	-0.016 (0.033)	(0.033)
Openness								
Desire to know	0.072 (0.032)	0.707 (0.036)	-0.055 (0.025)	(0.033)	0.036 (0.033)	0.666 (0.045)	-0.055 (0.034)	(0.034)
Originality	-0.019 (0.028)	0.711 (0.036)	0.033 (0.028)	(0.036)	0.024 (0.036)	0.604 (0.052)	0.099 (0.049)	(0.049)
Intelligence	-0.002 (0.028)	0.713 (0.035)	0.032 (0.027)	(0.030)	0.030 (0.030)	0.713 (0.044)	-0.019 (0.032)	(0.032)
Extraversion								
Fondness for large groups	-0.133 (0.032)	0.040 (0.039)	0.640 (0.036)	(0.033)	-0.079 (0.033)	-0.061 (0.035)	0.716 (0.043)	(0.043)
Leadership	-0.017 (0.033)	0.193 (0.042)	0.582 (0.036)	(0.036)	-0.105 (0.033)	0.308 (0.056)	0.534 (0.051)	(0.051)
Popular with other children	0.080 (0.022)	-0.066 (0.016)	0.838 (0.035)	(0.035)	0.172 (0.033)	-0.023 (0.038)	0.690 (0.046)	(0.046)
Root mean square error of approximation (RMSEA)		0.049				0.047		
90 Percent C.I.		(0.029, 0.070)				(0.022, 0.072)		
Probability RMSEA <= .05		0.485				0.535		
Comparative fit index (CFI)		0.989				0.989		
Tucker-Lewis index (TLI)		0.968				0.967		
Standardized root-mean-square-residual (SRMR)		0.016				0.017		
Sample size		695				536		

Notes: Factor loadings are based on oblique quartimin rotation. Factor loadings relating factors to the main corresponding measures are in bold. Calculations are based on the Terman data.

Table A-5: ML CFA Results for the COE-1922 Model

	Males								Females							
	model A		model B		model C		model D		model A		model B		model C		model D	
	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error	Factor Loading	Standard Error
Conscientiousness (C)																
Prudence	0.623 (0.043)		0.678 (0.037)		0.638 (0.042)		0.616 (0.042)		0.554 (0.048)		0.588 (0.043)		0.565 (0.048)		0.556 (0.048)	
Conscientiousness	0.894 (0.036)		0.884 (0.035)		0.851 (0.034)		0.896 (0.035)		0.931 (0.041)		0.925 (0.040)		0.894 (0.038)		0.922 (0.039)	
Truthfulness	0.748 (0.037)		0.756 (0.036)		0.782 (0.035)		0.750 (0.036)		0.746 (0.042)		0.750 (0.041)		0.773 (0.039)		0.753 (0.040)	
Desire to know	0.065 (0.042)				0.045 (0.044)		0.068 (0.042)		0.003 (0.053)				-0.016 (0.056)		-0.001 (0.053)	
Popular with other children	0.186 (0.046)				0.241 (0.051)		0.183 (0.045)		0.295 (0.054)				0.341 (0.060)		0.314 (0.057)	
Freedom from vanity					0.588 (0.049)								0.596 (0.056)			
Sympathy							0.434 (0.044)								0.420 (0.052)	
Openness (O)																
Desire to know	0.715 (0.048)		0.738 (0.039)		0.731 (0.049)		0.715 (0.048)		0.677 (0.063)		0.657 (0.046)		0.696 (0.066)		0.681 (0.064)	
Originality	0.710 (0.039)		0.693 (0.039)		0.703 (0.039)		0.710 (0.039)		0.658 (0.047)		0.630 (0.047)		0.656 (0.046)		0.659 (0.047)	
Intelligence	0.720 (0.038)		0.721 (0.039)		0.713 (0.038)		0.720 (0.038)		0.708 (0.046)		0.734 (0.046)		0.698 (0.045)		0.708 (0.045)	
Prudence	0.107 (0.042)				0.093 (0.042)		0.109 (0.042)		0.075 (0.050)				0.061 (0.050)		0.070 (0.049)	
Leadership	0.214 (0.048)				0.211 (0.048)		0.213 (0.047)		0.329 (0.051)				0.331 (0.051)		0.325 (0.051)	
Popular with other children	-0.086 (0.061)				-0.152 (0.069)		-0.082 (0.060)		-0.001 (0.069)				-0.046 (0.075)		-0.021 (0.073)	
Freedom from vanity					-0.277 (0.052)								-0.168 (0.061)			
Sympathy							0.039 (0.047)								0.094 (0.059)	
Extraversion (E)																
Fondness for large groups	0.628 (0.042)		0.613 (0.040)		0.604 (0.043)		0.638 (0.042)		0.691 (0.052)		0.625 (0.049)		0.685 (0.052)		0.679 (0.052)	
Leadership	0.563 (0.042)		0.651 (0.042)		0.540 (0.043)		0.561 (0.041)		0.507 (0.048)		0.668 (0.053)		0.508 (0.048)		0.497 (0.048)	
Popular with other children	0.840 (0.050)		0.811 (0.043)		0.901 (0.055)		0.832 (0.049)		0.693 (0.060)		0.684 (0.052)		0.716 (0.065)		0.720 (0.063)	
Truthfulness	0.055 (0.033)				0.047 (0.033)		0.056 (0.033)		0.015 (0.038)				0.005 (0.038)		0.014 (0.038)	
Desire to know	-0.083 (0.039)				-0.082 (0.039)		-0.083 (0.039)		-0.083 (0.055)				-0.094 (0.056)		-0.088 (0.055)	
Freedom from vanity					0.200 (0.042)								0.012 (0.054)			
Sympathy							0.183 (0.041)								0.172 (0.052)	
Root mean sq. err. of approx. (RMSEA)	0.042		0.067		0.059		0.047		0.050		0.085		0.056		0.046	
90 Percent C.I.	(0.024, 0.060)		(0.053, 0.081)		(0.045, 0.074)		(0.032, 0.062)		(0.030, 0.071)		(0.070, 0.101)		(0.040, 0.073)		(0.028, 0.064)	
Probability RMSEA <= .05	0.736		0.020		0.130		0.611		0.462		0.000		0.246		0.628	
Comparative fit index (CFI)	0.989		0.960		0.973		0.983		0.983		0.930		0.974		0.982	
Tucker-Lewis index (TLI)	0.976		0.941		0.947		0.967		0.963		0.895		0.948		0.965	
Stand. root-mean-sq.-resid. (SRMR)	0.021		0.046		0.026		0.023		0.024		0.056		0.026		0.024	
Sample size	695		695		695		695		536		536		536		536	

Notes: Factor loadings are based on the confirmatory factor analysis. Factor loadings relating factors to the corresponding possible dedicated measures are in bold. Calculations are based on the Terman data.

Table A-6: Measures from 1922 and 1940 that Proxy Factors Comparable to Big Five (Minor Cross-Loadings from Model A not Shown to Highlight the Key Measures)

Conscientiousness (1922)	Extraversion (1922)
Prudence and forethought	Fondness for large groups
Conscientiousness	Leadership
Truthfulness	Popularity with other children
Neuroticism (1940)	Openness (1922)
Miserable	Desire to know
Touchy	Originality
Periods of loneliness	Intelligence
Lonely when with others	
Remorseful and regretful	Agreeableness (1940)
Lacks self-confidence	Easy to get along with
Worries about humiliating experiences	Avoids arguments
Emotionally unstable	Critical
Easily hurt	Tactful
Hard to be serene	Unfeeling
Moody	Domineering
Sensitive	Inflated opinion of self

Note: Measures for 1940 factors are based on EFA and CFA performed in a companion paper. See Web Appendix for [Savellyev and Tan \(2019\)](#) for details about 1940 measures.

B Supplementary Figures and Tables

The section below is composed of figures and tables that are mentioned in various parts of the main paper as supplementary. Each paragraph below shortly describes one or several tables or figures.

Evolution of School Enrollment Figure B-1 shows the evolution of school enrollment, which sharply decreases with age. By age 30, only about 5% of the Terman population are in school. As discussed in the main paper, this result is one argument for the low likelihood of a reverse causal interpretation of results in a model conditional on survival to age 30.

Comparison of Models with Varying Starting Age Table B-1 shows that results of the model are robust to starting age from about age 12 (year 1922) to age 30. One reason for similar coefficients is low mortality before age 30.

MPH Model of Mortality and Logit Models if Education Choice for Females Table B-2 represents results of the Cox model for females and supplements the corresponding table for males (see Table 4 of the main paper). Unlike for males, neither single nor joint tests can be rejected.

The Measurement System The measurement system is an integral part of the system of equations jointly estimated in this paper, which allows the identification of the effect of latent factors on education and the hazard of death. Table B-3 shows that the main loadings, measures for which are shown in bold, are all statistically significant at the 1% level and tend to be of substantial size, around 0.5–0.9 or higher.⁷ In contrast, the cross-loadings measures (not in bold) are substantially smaller than the main loadings and tend to be less precisely determined. This observation is important since we use measures associated with the main loadings to interpret the factor, while cross-loadings account for small correlations that we

⁷The first loading is not 1 since I used a standard alternative normalization: setting variance of all latent factors to 1.

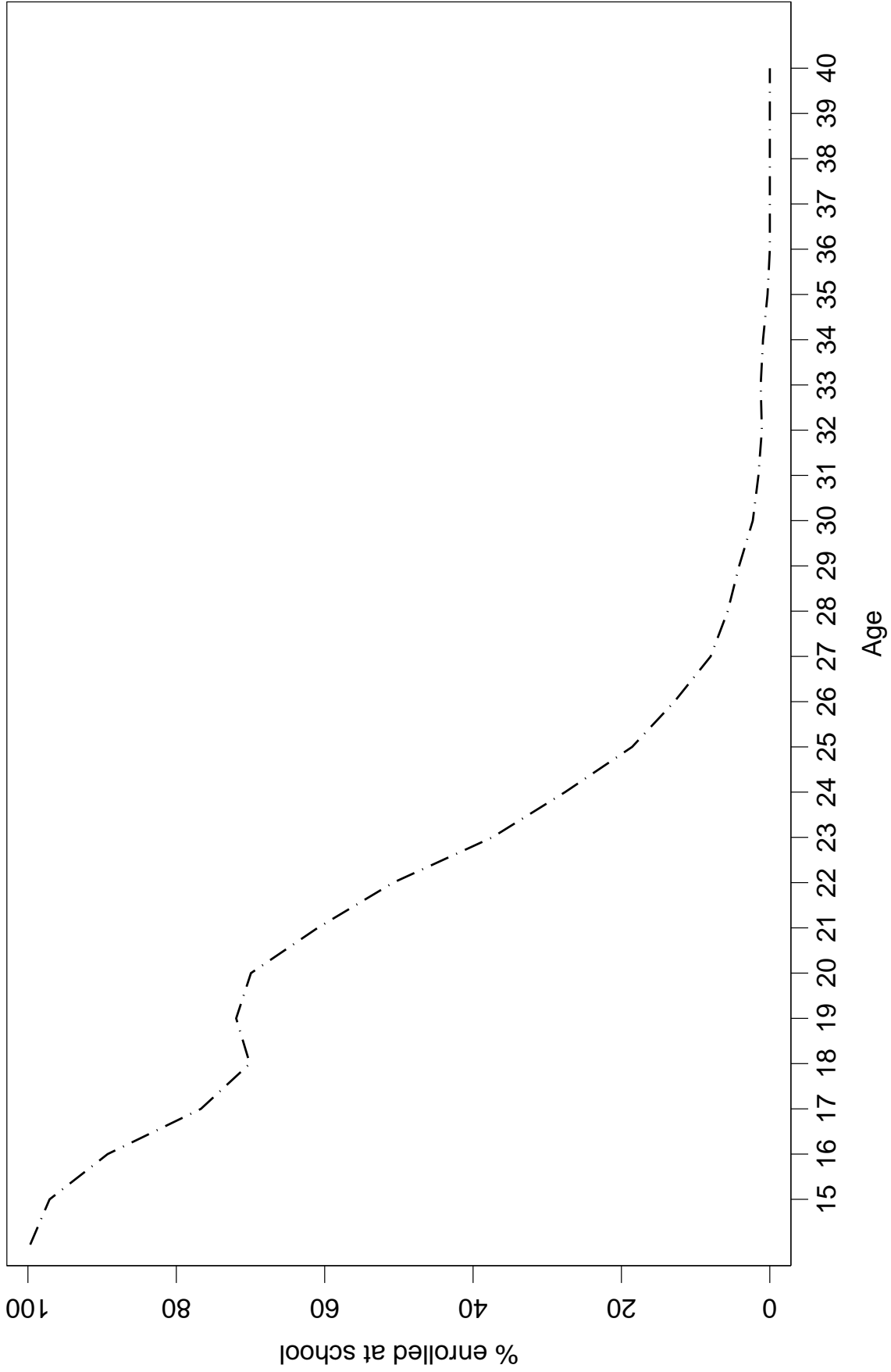
allow in order to make plausible exclusion restrictions, as discussed in detail in [Web Appendix A](#).

Checking Robustness of Estimates with Respect to Exclusion of Certain Groups from the Sample Table [B-4](#) shows the Cox model results for males for the full sample and various sub-samples. Sub-samples are the following: (1) violent deaths omitted; (2) suicides omitted; (3) auto crashes omitted; (4) alcohol-related deaths omitted; (5) all unknown causes of death omitted. We see that results are robust to excluding specific groups outlined above (there is some variation of estimates across cases, but results are qualitatively the same and numerically similar).

Robustness of the Main Model Based on 1922 Personality Measures to the Inclusion of Agreeableness and Neuroticism Factors Based on Measures from 1940 Table [B-5](#) compares a three-factor with a five-factor model. From the comparison we see that regression coefficients for Agreeableness and Neuroticism are small and statistically insignificant, both individually and jointly (the p -value for the Wald test is 0.960). We can also see that the omission of these two factors leads to only small changes in other regression coefficients. The omission of the two factors leads to a strong reduction in degrees of freedom and improvement in numerical stability of the model.

Robustness of the Main Full Likelihood Model to a Two-Step Estimation As discussed in the main paper, one common concern about using a full maximum likelihood model in research involving latent skills is that in the case of model misspecification, skills might be primarily driven by outcomes rather than the measurement system. Table [B-6](#) shows that this concern is not pertinent to results of this paper. For Model 2 of the table, I follow [Heckman et al. \(2018\)](#) and estimate only the measurement system in the first step and then estimate the maximum likelihood model while fixing loadings of the measurement system to the values estimated in the first step. The estimates of Model 2 are almost identical to those of Model 1 based on full maximum likelihood.

Figure B-1: Evolution of School Enrollment*



Notes: (*)Attending high school, undergraduate, or graduate school. Calculations are based on the Terman data.

Table B-1: Comparison of Models with Varying Starting Age, Males

	Age at Start (Year 1922)		Starting Age 23		Starting Age 25		Starting Age 27		Starting Age 30	
	Mortality MPH (1)	Education Ordered (2)	Mortality MPH (1)	Education Ordered (2)	Mortality MPH (1)	Education Ordered (2)	Mortality MPH (1)	Education Ordered (2)	Mortality MPH (1)	Education Ordered (2)
Education										
High school degree	0.731 *** (0.183)	-	0.712 *** (0.185)	-	0.737 *** (0.186)	-	0.752 *** (0.187)	-	0.774 *** (0.188)	-
Some college education	0.452 *** (0.141)	-	0.394 *** (0.143)	-	0.389 *** (0.144)	-	0.387 *** (0.144)	-	0.394 *** (0.144)	-
Bachelor's degree or above	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
Skills										
Conscientiousness	-0.138 * (0.073)	0.199 (0.126)	-0.145 * (0.075)	0.204 (0.129)	-0.141 * (0.075)	0.205 (0.128)	-0.130 * (0.074)	0.218 * (0.128)	-0.129 * (0.073)	0.210 * (0.127)
Extraversion	-0.122 * (0.070)	-0.169 (0.123)	-0.123 * (0.071)	-0.166 (0.124)	-0.123 * (0.071)	-0.167 (0.124)	-0.128 * (0.071)	-0.172 (0.125)	-0.138 * (0.071)	-0.163 (0.125)
Openness	0.114 (0.081)	0.167 (0.145)	0.096 (0.082)	0.184 (0.150)	0.090 (0.082)	0.182 (0.148)	0.093 (0.083)	0.175 (0.148)	0.093 (0.082)	0.178 (0.147)
IQ	-0.050 (0.062)	0.298 *** (0.103)	-0.034 (0.062)	0.295 *** (0.103)	-0.028 (0.062)	0.298 *** (0.103)	-0.027 (0.063)	0.297 *** (0.103)	-0.025 (0.062)	0.299 *** (0.105)
Observations	711		706		702		700		695	

Notes: Model coefficients are presented with standard errors shown in parentheses. For the purpose of numerically-stable robustness check, I condition all models on background controls but not on unobserved heterogeneity. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level.

Table B-2: MPH Model of Mortality and Logit Models of Education Choice, Females

Outcome Sub-Model Type	Model 1		Model 2		Model 3		Model 4		Model 5		PH-test
	Mortality MPH (1)	Education Ordered logit (2)	Mortality MPH (3)	Education Ordered logit (4)	Mortality MPH (5)	Education Ordered logit (6)	Mortality MPH (7)	Mortality MPH (8)	Mortality MPH (8)		
High school degree	-0.042 (0.271)	-	-0.042 (0.268)	-	-0.02 (0.269)	-	-0.01 (0.270)	-0.003 (0.240)	0.589		
Some college education	-0.156 (0.209)	-	-0.14 (0.200)	-	-0.111 (0.197)	-	-0.103 (0.195)	-0.039 (0.184)	0.629		
Bachelor's degree or above	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted		
Conscientiousness	0.153 (0.109)	-0.226 (0.162)	-0.139 (0.102)	0.208 (0.148)	-	-	-	-	0.693		
Extraversion	0.057 (0.113)	-0.188 (0.153)	0.003 (0.103)	-0.189 (0.145)	-	-	-	-	0.876		
Openness	-0.119 (0.134)	-0.056 (0.185)	0.124 (0.125)	0.067 (0.175)	-	-	-	-	0.175		
IQ	-0.018 (0.101)	0.103 (0.124)	-0.098 (0.094)	0.112 (0.115)	-0.080 (0.087)	0.163 (0.114)	-	-	0.732		
Joint test p -values	0.8238		0.6881		0.8522		0.8655	0.9770	0.8597		
Unobs. Heterogeneity (τ)	Yes	No	No	No	No	No	No	No	No		
Background controls (X)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No		

Notes: Model coefficients are presented with standard errors shown in parentheses. The table shows the same model as Table 4 of the main paper, but for females. Results are based on the Terman data.

Table B-3: The Measurement System

	Males		Females	
	coef- ficient	standard error	coef- ficient	standard error
Conscientiousness				
Prudence and forethought	0.637 ***	0.047	0.569 ***	0.053
Conscientiousness	0.892 ***	0.031	0.909 ***	0.040
Truthfulness	0.752 ***	0.039	0.771 ***	0.044
Fondness for large groups	-0.127 **	0.056	0.096	0.074
Desire to know	0.063	0.046	-0.038	0.063
Popularity with other children	0.106 *	0.061	0.399 ***	0.090
Extraversion				
Fondness for large groups	0.642 ***	0.044	0.695 ***	0.056
Leadership	0.586 ***	0.047	0.485 ***	0.052
Popularity with other children	0.872 ***	0.056	0.727 ***	0.068
Truthfulness	0.054	0.035	0.002	0.043
Desire to know	-0.093 **	0.045	-0.105	0.066
Openness				
Desire to know	0.740 ***	0.056	0.705 ***	0.065
Originality	0.705 ***	0.039	0.663 ***	0.049
Intelligence	0.721 ***	0.042	0.703 ***	0.052
Prudence and forethought	0.099 **	0.047	0.060	0.054
Leadership	0.136 **	0.063	0.363 **	0.066
Popularity with other children	-0.155 **	0.071	-0.030	0.083
Background Variables	Yes		Yes	
Sample	695		536	

Notes: Measures corresponding to diagonal loadings are in bold; measures that serve as cross-loadings are not bolded. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level.

Table B-4: Checking Robustness of Estimates with Respect to Exclusion of Groups with Specific Causes of Deaths from the Sample, Males

	full sample	sub- sample 1	sub- sample 2	sub- sample 3	sub- sample 4	sub- sample 5
	main model	all violent deaths omitted	suicide omitted	auto crashes omitted	alcohol- related deaths omitted	cause of death unknown omitted
Education						
High school	0.974 *** (0.215)	0.998 *** (0.227)	0.976 *** (0.226)	0.984 *** (0.214)	0.988 *** (0.211)	0.984 *** (0.264)
Some college	0.525 *** (0.167)	0.475 *** (0.171)	0.495 *** (0.171)	0.530 *** (0.166)	0.530 *** (0.166)	0.533 *** (0.189)
Bachelor's, Master's or Doctorate degree	-	-	-	-	-	-
Skills						
Conscientiousness	-0.160 * (0.086)	-0.166 * (0.088)	-0.153 * (0.087)	-0.170 * (0.087)	-0.170 * (0.087)	-0.198 ** (0.100)
Extraversion	-0.158 ** (0.076)	-0.146 * (0.079)	-0.161 ** (0.079)	-0.149 * (0.077)	-0.161 ** (0.077)	-0.188 ** (0.085)
Openness	0.095 (0.088)	0.085 (0.087)	0.087 (0.087)	0.100 (0.088)	0.108 (0.089)	0.129 (0.099)
IQ	0.040 (0.069)	0.054 (0.072)	0.040 (0.071)	0.050 (0.069)	0.048 (0.069)	0.047 (0.075)
Background Variables	Yes	Yes	Yes	Yes	Yes	Yes
Joint test p -value	0.000	0.000	0.000	0.000	0.000	0.000
Sample size	695	670	677	692	690	623

Notes: Model coefficients are presented with standard errors shown in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Joint test is for all education and skill variables.

Table B-5: Robustness of the Main Model Based on 1922 Personality Measures to the Inclusion of Agreeableness and Neuroticism Factors Based on Measures from 1940

Outcome Sub-model type	Model 1 (main model)		Model 2 (Robustness Check)	
	Mortality MPH (1)	Education ordered logit (2)	Mortality MPH (3)	Education ordered logit (4)
High school degree	0.974 *** (0.215)	–	0.990 *** (0.214)	–
Some college education	0.525 *** (0.167)	–	0.515 *** (0.167)	–
Bachelor's degree or above	omitted		omitted	
Conscientiousness	-0.160 * (0.086)	0.192 (0.141)	-0.152 * (0.088)	0.210 (0.148)
Extraversion	-0.158 ** (0.076)	-0.195 (0.135)	-0.145 * (0.078)	-0.187 (0.137)
Openness	0.095 (0.088)	0.155 (0.157)	0.087 (0.089)	0.142 (0.164)
Agreeableness	–	–	-0.071 (0.076)	-0.083 (0.150)
Neuroticism	–	–	-0.055 (0.070)	-0.037 (0.125)
IQ	0.040 (0.069)	0.399 *** (0.135)	0.045 (0.070)	0.399 *** (0.137)
Joint test p -values	0.0000		0.0000	
Wald Test for A and N	–		0.9598	
Unobs. Heterogeneity (τ)	Yes		Yes	
Background controls (X)	Yes		Yes	

Notes: Model coefficients are presented with standard errors shown in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Sample size is 695.

Table B-6: Robustness of the Main Full Likelihood Model to a Two-Step Estimation (First Measurement System, then Outcome Equations)

Outcome	Model 1 (main model)		Model 2 (Robustness Check)	
	Mortality	Education	Mortality	Education
Sub-model type	MPH	ordered logit	MPH	ordered logit
	(1)	(2)	(3)	(4)
High school degree	0.974 *** (0.215)	–	0.971 *** (0.215)	–
Some college education	0.525 *** (0.167)	–	0.523 *** (0.167)	–
Bachelor's degree or above	omitted		omitted	
Conscientiousness	-0.160 * (0.086)	0.192 (0.141)	-0.157 * (0.082)	0.196 (0.136)
Extraversion	-0.158 ** (0.076)	-0.195 (0.135)	-0.155 ** (0.074)	-0.193 (0.133)
Openness	0.095 (0.088)	0.155 (0.157)	0.089 (0.082)	0.145 (0.151)
IQ	0.040 (0.069)	0.399 *** (0.135)	0.040 (0.069)	0.400 *** (0.135)
Joint test <i>p</i> -values	0.0000		0.0000	
Unobs. Heterogeneity (τ)	Yes		Yes	
Background controls (X)	Yes		Yes	

Notes: Model coefficients are presented with standard errors shown in parentheses. Asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level. Sample size is 695.

C The Likelihood Function

This section provides the likelihood function for the main model corresponding to equations (1–4) of the main paper. I use the expectation-maximization (EM) algorithm for estimation, as the integral presented below has no closed form solution. For an example of a comparable likelihood function see, for instance, [Larsen \(2004\)](#).

$$\begin{aligned} \mathcal{L} = & \prod_{i=1}^N \int \left\{ \left[\sum_{k=1}^K \pi_{ki} \cdot p_{1ik}^{z_{i1}} \cdot p_{2ik}^{z_{i2}} \cdot p_{3ik}^{z_{i3}} \right. \right. \\ & \times \left. \left. \{ \lambda_0(t_i) \exp(g_\lambda + \mu_{\lambda k}) \}^{\delta_i} \exp \left\{ - \exp(g_\lambda + \mu_{\lambda k}) \int_0^{t_i} \lambda_0(s) ds \right\} \right] \right. \\ & \times \left. \prod_{r=1}^R \prod_{q_r=1}^{Q_r} \frac{1}{\sigma_{rq_r}} \phi \left(\frac{M_{rq_r} - g_{rq_r}}{\sigma_{rq_r}} \right) \right\} dF_{\Theta}(\boldsymbol{\theta}), \end{aligned} \quad (\text{C.1})$$

where

- i is an index for individuals
- N is the estimation sample size
- Θ is a random vector of continuous personality factors and $\boldsymbol{\theta}$ is its realization
- $\pi_{ji} = \frac{\exp(\mathbf{c}_{1k}\boldsymbol{\theta}_i + \mathbf{c}_{2k}IQ_i)}{\sum_{j=1}^K \exp(\mathbf{c}_{1j}\boldsymbol{\theta}_i + \mathbf{c}_{2j}IQ_i)}$, $k = 1, \dots, K$ is the probability that individual i belongs to latent class k
- $p_{1ik} = 1 - F(g_{Di} + \mu_{D1k})$, $p_{2ik} = F(g_{Di} + \mu_{D1k}) - F(g_{Di} + \mu_{D2k})$, and $p_{3ik} = F(g_{Di} + \mu_{D2k})$ are probabilities of ordered educational outcomes, where $g_{Di} = \mathbf{b}_1\boldsymbol{\theta}_i + b_2IQ_i + \mathbf{b}_3\mathbf{x}_i$, IQ_i is the intelligence measurement, \mathbf{x}_i is a vector of background variables, and μ_{D1k} and μ_{D2k} are ordered logit's thresholds that differ by latent class k .
- $z_{ij} = 1$ if $D_i = j$ and zero otherwise, $j = 1, 2, 3$
- t_i is time to either death or right censoring conditional to survival to time t_0 and using t_0 as the beginning of time

- $\lambda_0(t_i)$ is the baseline hazard function estimated nonparametrically (like in Cox model)
- δ_i is the right-censoring indicator: $\delta_i = 1$ if individual i is uncensored and zero if the individual is right-censored
- $g_\lambda = \mathbf{b}_1\boldsymbol{\theta}_i + b_2IQ_i + b_3D_{1i} + b_4D_{2i} + b_5\mathbf{x}_i$
- $\mu_{\lambda k}$ is the intercept that is allowed to vary by latent class k (the intercept is normalized to zero for the base class, which is necessary for identification)
- r is the index for the latent skill type, and R is the total number of distinct latent skills in the model
- q_r is the index for distinct measures M_{rq_r} of latent class of type r , with the total number of measures of type r being Q_r
- $\sigma_{rq_r}^2$ is the residual variance of measure M_{rq_r}
- $g_{rq_r} = \boldsymbol{\alpha}_{rq_r}\boldsymbol{\theta}_i$, where $\boldsymbol{\alpha}_{rq_r}\boldsymbol{\theta}_i$ is a linear combination of latent factors loaded on measure M_{rq_r} , weighted with a vector or factor loadings $\boldsymbol{\alpha}_{rq_r}$. Some measures are loaded on only a subset of factors, as described in Web Appendix A, in which case the elements of $\boldsymbol{\alpha}_{rq_r}$ corresponding to irrelevant factors are set to zero

References

- Bentler, P. M. and D. G. Bonett (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88(3), 588–606.
- Bollen, K. and J. Long (1993). *Testing Structural Equation Models*. SAGE Focus Editions. SAGE Publications.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research* 1(2), 245–276.
- Du, T. J. L. and J. S. Tanaka (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology* 74(74), 625–635.
- Friedman, H. S., J. S. Tucker, J. E. Schwartz, L. R. Martin, C. Tomlinson-Keasey, D. L. Wingard, and M. H. Criqui (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology* 68(4), 696–703.
- Friedman, H. S., J. S. Tucker, C. Tomlinson-Keasey, J. E. Schwartz, D. L. Wingard, and M. H. Criqui (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology* 65(1), 176–185.
- Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika* 19(2), 149–161.
- Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health and smoking. *Journal of Political Economy* 126, 197–246.
- Heckman, J. J., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms

- through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–2086.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30(2), 179–185.
- John, O. P. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2 ed.), Chapter 4, pp. 102–138. New York: The Guilford Press.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20(1), 141–151.
- Kaiser, H. F. (1961). A note on Guttman’s lower bound for the number of common factors. *British Journal of Statistical Psychology* 14(1), 1–2.
- Kline, R. (2011). *Principles and Practice of Structural Equation Modeling*. Methodology in the social sciences. Guilford Press.
- Kolenikov, S. (2009, September). Confirmatory factor analysis using confa. *Stata Journal* 9(3), 329–373.
- Larsen, K. (2004). Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes. *Biometrics* 60(1), 85–92.
- Martin, L. R. and H. S. Friedman (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality* 68(1), 85–110.
- Martin, L. R., H. S. Friedman, and J. E. Schwartz (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology* 26(4), 428–436.

- Miles, J. and M. Shevlin (2007). A time and a place for incremental fit indices. *Personality and Individual Differences* 42, 869–874.
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* 77(5), 1447–1479.
- Roberts, B. W., J. J. Jackson, J. V. Fayard, G. Edmonds, and J. Meints (2009). *Handbook of Individual Differences in Social Behavior*, Chapter Conscientiousness, pp. 369–381. New York: Guilford.
- Savelyev, P. A. and K. Tan (2019). Socioemotional skills, education and health-related outcomes of high-ability individuals. *American Journal of Health Economics* 5(1), 250–280.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences* 42, 893–898.
- Terman, L. M. and R. R. Sears (2002). *The Terman Life-Cycle Study of Children with High Ability, 1922–1986*, Volume 1, 1922–1928. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika* 41(3), 321–327.
- Zwick, W. R. and W. F. Velicer (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* 99(3), 432–442.