

*Comparing and Validating Measures of
Character Skills: Findings from a
Nationally Representative Sample*

*Gema Zamarro, Albert Cheng,
M. Danish Shakeel, Collin Hitt*

Paper No: 2016-011

**CESR-SCHAEFFER
WORKING PAPER SERIES**

The Working Papers in this series have not undergone peer review or been edited by USC. The series is intended to make results of CESR and Schaeffer Center research widely available, in preliminary form, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. CESR-Schaeffer working papers can be cited without permission of the author so long as the source is clearly referred to as a CESR-Schaeffer working paper.

Comparing and Validating Measures of Character Skills:
Findings from a Nationally Representative Sample

Gema Zamarro*

Albert Cheng

M. Danish Shakeel

Collin Hitt

University of Arkansas

May, 2016

* Corresponding author. University of Arkansas. Address: 219B Graduate Education Building. College of Education and Health Professions Fayetteville, AR, USA 72701. Phone: 479-575-7024. Email: gzamarro@uark.edu. We thank Bart Orriens for all his help with data collection. We also thank Lindsay Weixler and other seminar participants at the Association of Education Finance and Policy's 41st Annual Conference and at the University of Arkansas Department of Education Reform Brownbag series for their comments on earlier versions of this paper. All remaining errors are our own.

Abstract

Though researchers now are aware of the potential importance of character skills, such as conscientiousness, grit, self-control, and a growth mindset, researchers struggle to find reliable measures of these skills. In this paper, we use data collected from the Understanding America Study, a nationally representative internet panel to study the validity of innovative measures of character skills based on measures of survey effort. We believe surveys themselves can be seen as a behavioral tasks and that respondents provide meaningful information about their character skills by way of the effort they put forward on surveys. In particular, we compare measures of grit, conscientiousness and other personality traits, and growth mindset, based on self-reports, and survey effort measures of character. We study the relationship across each other and their relationship with academic and life outcomes such as income and labor-market outcomes, after controlling for cognitive ability and other relevant demographic characteristics. Our results show that survey effort measures of character skills, in particular measures of careless answering in surveys, show great promise for being good proxy measures of relevant non-cognitive skills.

Keywords: character skills, grit, conscientiousness, behavioral tasks, survey effort.

JEL codes: C83, C91

I. Introduction

It is well established in the human capital literature that IQ or cognitive ability is an important determinant of educational and labor-market outcomes. For example, individuals with higher cognitive ability are found to exhibit higher educational attainment levels and earnings (Becker 1964; Hanushek & Woessmann 2012). More recently, human capital research is recognizing the crucial role that so-called noncognitive skills or character skills¹ play in realizing the same outcomes, even after controlling for IQ and cognitive ability. Character skills such as grit, conscientiousness, self-control, and growth mindset have also been found to independently influence health outcomes and the propensity to engage in criminal behavior (Almlund et al., 2011; ter Weel, 2008; Heckman, Stixrud, & Urzua 2006).

Despite a growing body of research indicating the importance of character skills, evaluations of educational policy and interventions intended to improve human capital rarely analyze impacts on these skills. This is not a trivial issue, overlooking the possibility that these policies or interventions impacts character skills may lead to an incomplete appraisal of the ways in which they benefit students or other stakeholders. For example, evaluations of early childhood education programs, charter schools, and private school vouchers have often found little to no impacts on student achievement and other indicators of cognitive ability. Yet these same evaluations document large gains in educational attainment, employment income, and health outcomes as well as reductions in criminal behavior (Sass et al., 2013; Campbell et al. 2014; Chingos & Peterson, 2015; Cowen et al., 2013; Heckman et al. 2010; Heckman et al, 2013; Wolf et al., 2013). Researchers speculate that these differences in longer-run outcomes, despite no noticeable differences in achievement, are due to these programs effects on their participants' character skills. Had evaluations of these programs been conducted by only considering impacts on cognitive ability, their benefits would have been understated.

Despite the salience of character skills, it is difficult to measure character skills with validity. One way to measure character skills would be to use proxies that are commonly available in administrative data. Information about student grades, attendance, and behavioral

¹ We use the term character skills throughout the article, while recognizing the use of other terms, such as noncognitive skills, soft skills, or social-emotional skills, to refer to the same or similar concepts.

reports have been used to this end (Gershenson, 2016; Jackson, 2012). However, this information is seldom available for researchers. As a result, researchers mostly rely on data from self-reported psychometric scales, where study participants are asked to answer a series of Likert-type items. Although relatively easy to collect, these type of self-reported measures have been shown to be affected by social desirability bias, reference group bias, and other threats to validity (Dobbie & Fryer, 2015; West et al., 2016). Even if these sources of bias were absent from self-reported measures, these measures are rarely, if at all, validated in nationally representative samples. Self-reported measures have typically been validated in convenience samples, calling into question the extent to which they can be used for more general populations.

An alternative approach to self-reported measures of character skills is to obtain measures based on performance tasks. In a performance task, a researcher asks respondents to complete a carefully-designed task and interprets their behavior as some indication of an underlying character skill. Variation in respondent behaviors provides meaningful differences in behaviors as indicative of their level of a given skill. Mischel and Ebbesen's (1970) famous "Marshmallow Test", where young children were presented with the options of eating one marshmallow or waiting until the experimenter returned (about 15 minutes) to get two marshmallows instead, is an example of a performance task designed to measure self-control. Though performance-task measures do not suffer from the same sources of biases than self-reports, they have limitations of their own. For instance, it is not always clear that artificial tasks completed in a lab setting are generalizable to other contexts. Nor is it clear that behavioral tasks capture the non-cognitive skill that it purports to capture. Tasks are generally very costly and difficult to collect in large samples and they are difficult to be implemented multiple times to study progress, as participants might show learning effects after having performed the task once (Duckworth & Yeager, 2015; Heckman & Falk, 2009).

These limitations have generated calls to improve measurement of character skills (see, Duckworth and Yeager, 2015). This paper is motivated by this call. We use data from the Understanding America Study (UAS), a nationally-representative internet panel of US adults maintained by the University of Southern California, to study the potential of a new performance-task measure to capture underlying character skills. We refer to these measures as survey-effort measures of character skills. More specifically, we argue that survey questionnaires can be seen as tasks that require effort to complete and that respondents reveal something about

their character skills through the effort that they exhibit to complete the questionnaires. We study the possibilities of two survey-effort measures: Item non-response rates and careless answering. Previous research has found that these measures in adolescents are predictive of later educational attainment, independent of cognitive ability (Hitt, Trivitt, Cheng (forthcoming) & Hitt (2015)).

In this paper, we provide additional validation of survey-effort measures of non-cognitive skills by examining their relationship with self-reported measures of the Big Five personality traits, grit and growth mindset, and their association with education, income and labor-market outcomes, after controlling for cognitive ability and other relevant demographic characteristics. This paper is the first to simultaneously explore the relationships between survey-effort measures of character skills and self-reported measures as no previous data set possess all these measures together. Moreover, this paper represents the first attempt to conduct such a validation exercise in a nationally representative sample. Similar validation studies of character-skill measures usually rely on small samples of convenience.

Our results show that survey-effort measures of character skills, in particular measures of careless answering, show great promise for being valid proxies of important character skills. Careless answering correlates mostly with self-reported measures of conscientiousness and neuroticism, just as self-reported grit does although self-reported grit also correlates with other personality traits. Careless answering presents stronger correlations with education and labor-market outcomes than self-reported grit. In addition, a significant advantage of careless answering and, more generally, survey-effort measures is that they are not affected by social desirability bias, reference group bias, and other threats to validity that are unique to self-reported measures.

The remainder of this paper is divided into four sections. In section 2, we describe the Understanding America Study, the source of data for this paper. In this section, we also discuss in greater detail the character skills that we aim to measure and how researchers have studied them in the past. At this point, we also describe our proposed alternative measures of character skills based on measures of survey effort. Next, in section 3, we describe the methods that we will use to validate the measures of character skills in our data. In section 4, we present the results of our analysis. Finally, in section 5, we discuss these results and offer concluding remarks about these measures of character skills and their place in future research.

2. Data and Measures

For this paper, we use data collected in the Understanding America Study (UAS). The UAS is a new household panel recruited by the University of Southern California, comprising a nationally-representative sample of approximately 2,270 US households.² UAS respondents complete up to 30-minute surveys in waves that occur once or twice per month³. Respondents receive compensation for their time spent answering questions at a rate of \$20 per 30 minutes of interview time. Annual attrition rates are modest (on the order of 6 percent per year).

Our results are based on a survey wave that we designed. This wave includes self-reported grit and growth mindset measures, along with detailed information on school and family experiences during the respondents' childhood. We collected data for 1,729 respondents between June 2015 and January 2016.

A key feature of the panel structure of the UAS is the ability to link data across survey waves. In this paper, we use this unique feature and link our collected data to different waves containing information about the respondents' work status, educational background, cognitive ability, and personality traits (e.g., Big Five Inventory [John and Srivastava, 1999]). We describe this information in greater detail in the next sections. We also exploit the panel feature of the data to create our survey-effort measures of character skills that will be based on response patterns observed over multiple waves of data and constructed using waves different from the ones containing information for the analysis. Finally, sample weights are also provided to maintain representativeness of the U.S population.

Our primary aim is to validate innovative survey-effort measures of character skills by studying their correlations with self-reported measures and comparing their predictive power of long-run life outcomes. In particular, we focus on comparisons of measures that we believe to capture dimensions of conscientiousness, self-control and grit. What follows is a description of these character skills, how they are collected in our survey and used in the analysis, and how prior research has validated them.

2.1 Self-Reported Measures

² For more information visit: http://static.usc.edu/data_toolbox/understanding_america_study

³ It is important to note that participants are not limited to households who have computer hardware or purchase internet access. The UAS research team provides internet access and hardware, such as tablets, so that all households in the sample may participate.

Self-reported measures of conscientiousness and other character skills are widely available and frequently used by researchers. Administering a survey instrument containing this type of measure is a convenient way to obtain information about the respondents' character skills. Many of these instruments are easily accessible and have undergone empirical testing of their validity. However, they are often validated using convenience samples that are readily available to researchers. This limitation raises the issue of whether the scales are valid for different samples or segments of the general population. Our unique dataset allows us to validate these measures in a nationally representative sample of adults. Our work, then, represents one of the first efforts to assess the validity of these measures, many of which are only recently developed, in a nationally representative sample.

Grit

Duckworth et al. (2007) define grit as “perseverance and passion for long-term goals” (p. 1087). Our measure of grit is based on the eight-item grit scale developed by Duckworth and Quinn (2009). This scale asks respondents to answer 8 questions about themselves on a 5-point scale (Very much like me; Mostly like me; Somewhat like me; Not much like me; Not like me at all). Statements include, among others, “I am a hard worker”, “I am diligent”, and “Setbacks don't discourage me”. Averaging the scores from responses to each item, a grit score is computed for each respondent.

Duckworth & Quinn (2009) have shown that grit scores predict retention rates among West Point first-year cadets and success among National Spelling Bee participants. They also show that grit is positively correlated with GPAs (independently from IQ) among adolescents, as well as, educational attainment and career stability among adults. Although these results provide some assurance that the grit scale measures what it claims to measure, all of these validation samples are nonrandom. We address this limitation by validating grit in a nationally representative sample.

The Big Five Personality Traits

The Big Five is a taxonomy to understand personality traits. Under this framework, salient personality traits fall into one of five broad dimensions: conscientiousness, agreeableness, neuroticism, extroversion, and openness (Funder, 2001; John 1990; John, Naumann, & Soto,

2008). Other researchers have established subcategories within each of these five dimensions. For instance, traits such as orderliness, industriousness, responsibility and self-control are found to be separate facets of conscientiousness (Jackson et al., 2010; MacCann, Duckworth, & Roberts, 2009). Our measures of the Big Five personality traits are based on a 44-item scale developed by John, Donahue, and Kentle (1991). Respondents receive a score from 1 to 5 on each of the five dimensions represented in the scale based on their responses to each of the 44 items. The Big Five framework is widely used in the study of personality psychology and, more recently, economics (Almlund et al., 2011; Borghans et al., 2008). Like the grit scale, self-reported measures of the Big Five personality traits have been validated in numerous samples of convenience (See, e.g. John, Naumann, & Soto, 2008; John & Srivastava, 1999).

Growth Mindset

In addition to grit and the Big Five personality traits, we also study the properties of self-reported measures of growth mindset. Growth mindset is the extent to which one believes that intelligence is not fixed but malleable. People with the opposite perspective, a fixed mindset, believe that intelligence is fixed and cannot be changed. Prior work demonstrates that students who more often exhibit a growth mindset than a fixed mindset tend to experience better grades, standardized test scores, and other academic outcomes (Grant & Dweck 2003; Blackwell, Trzesniewski, & Dweck, 2007).

Our measure of growth mindset is based on two questions previously used in other large-scale longitudinal data sets (e.g., Ingels et al. 2004). Specifically, respondents are asked to indicate their level of agreement on a five-point scale with the following two statements: (a) Most people can learn to be good at math and (b) You have to be born with the ability to be good at math.

2.2 Survey-effort Measures of Character Skills

Although easy to collect, self-reported measures of non-cognitive skills are known to be potentially affected by significant biases. Consider, for example, the grit scale which has an item that asks a respondent to indicate how much they agree with the statement: “I am a hard worker.” Respondents answering this question must rely upon their own idiosyncratic standards of what it means to be a hard worker. Importantly, responses provided based upon different standards are no longer comparable. This limitation is a particularly acute in program evaluation as

interventions may not only alter the level of a specific character skill like grit but also the internal standards that respondents rely upon to complete the questions on the grit scale. This problem is referred to as reference-group bias and has been proposed to explain counterintuitive results in experimental evaluations of charter schools. According to this research, attending these charter schools yielded outcomes such as higher test scores, lower rates of teen pregnancy, and lower criminal behavior. Despite these results, and other work indicating a positive relationship between grit and these outcomes, participants attending these charter schools reported having lower levels of grit than their counterparts attending traditional public schools (Dobbie & Fryer, 2015; West et al., 2016).

Other sources of bias, such as social desirability bias, may be present as well. It is possible that most respondents will indicate that they are hard workers and that very few respondents will disagree with the proposition that they are not. Indeed, in the UAS, about 92% of the 1,729 respondents reported they agreed or strongly agreed with the grit scale item statement “I am a hard worker”. Only 8% of 1,729 respondents reported to be neutral, disagree or strongly disagree. Similar responses were observed for the statement “I am diligent” where 84% of respondents reported they agreed or strongly agreed. Although it is not clear that such skewed distributions distort or add noise to self-reported measures of grit, they do give reason to pause and ponder how respondents are answering these items.

More important, these issues could impair program evaluations. Reference-group bias, for instance, could lead to incorrect conclusions. Research of interventions that truly improve character skills while also raising standards by which participants assess themselves on having those character skills may show negative impacts on those skills. Similarly, social desirability bias, especially if its effects are not constant across different groups of the population, could lead to biased conclusions on the effects of interventions. Moreover, social desirability bias also reduces study power if it leads to little variation in how participants respond to the survey items. An alternative way to measure character skills that would not be affected by these biases is to use performance-task based measures.

In a performance-task measure, respondents are asked to perform a specific, carefully-designed task. Meaningful differences in behaviors are interpreted to indicate differences in the level of a given character skill. Performance-task measures are not as extensively affected by the same limitations as self-reported measures. Respondents are typically unaware that researchers

are assessing them on the amount of effort they put into completing surveys, alleviating issues related to social desirability bias.⁴ Furthermore, performance tasks reveal a respondent's behavior without asking them to appeal to their own internal standards to assess themselves on character skills. This feature of performance tasks helps to address issues related to reference-group bias.

We argue that working on a survey can be viewed as performance tasks and that parameterizations of survey effort can lead to meaningful measures of character skills. Survey effort can be measured by analyzing response patterns within surveys. Recent evidence has highlighted the potential of studying response patterns as a way of quantifying and recovering latent information about a respondents' character skills (Hitt, Trivitt, & Cheng, forthcoming, Hitt, 2015; see also Borghans & Schils, 2015; Mendez, Zamorro, Clavel & Hitt, 2015). More specifically, completing a survey presumably requires diligence and sustained effort. We submit that the extent to which the respondent leaves items blank is a reflection of a lack of character skills associated with conscientiousness. Other respondents lacking conscientiousness may not skip items but hastily rush through a survey, providing thoughtless and, therefore, inconsistent answers. We parameterize these two respective behaviors and refer to them as survey-effort measures of conscientiousness. In the next section, we describe how we build these measures: (1) item nonresponse rate and (2) careless answering.

2.2.1 Item Non-response Rates

We first use survey item non-response rates as a behavioral task measure of conscientiousness. Item non-response rates are defined as the percentage of items that respondents skipped out of the total number of items they were required to complete in a given survey. We compute the item non-response rates for surveys in five waves of data in the UAS⁵ that were particularly longer in length and so, presented more potential for observing patterns of item non-response. We then take the average item non-response rate across waves and within each individual as an indication of the level of each respondent's conscientiousness when completing surveys.

⁴ Of course, it is possible that these measures could eventually become distorted if they are repeatedly used, say, in high-stakes settings and respondents discover they are being evaluated on it.

⁵ The UAS survey waves used for this measure included: UAS11, UAS12, UAS20, UAS21 and UAS22.

Item non-response rate is calculated based upon data that are not used to create other variables for analysis. In particular, we omit information from the first wave of UAS in this average. The first wave of UAS includes information on many self-reported measures of character skills we use in the analysis as well as measures of cognitive ability. Because we desire to correlate individual average item response rates to self-reported measures of character skills included in this first wave in our validation exercise, we omit the first wave of UAS to avoid confounding variation between the two. In fact, item response rates were very high for this first survey, with an average item response rate of 98 percent. Respondents likely exhibited more diligence as this was the very first survey they took in the UAS, assuaging concerns that self-reported measures of character skills and measures of cognitive ability are distorted by low survey effort. This feature strengthens the validation exercise.

Hitt, Trivitt, and Cheng (forthcoming) have validated item non-response rates as a proxy for conscientiousness among school age children. They demonstrate that the extent to which adolescents skip items on paper-pencil surveys performed in school, is predictive of their educational attainment by adulthood. This relationship persists independently of measures of cognitive ability, and the authors find some evidence that item non-response rates appears to capture persistence and effort. Nonetheless, additional validation of item non-response rates as a measure of conscientiousness would be worthwhile, especially in the population of adults and with a different mode of interview, in this case an internet survey. We attempt to do exactly this analysis in this article.

2.2.2 Careless Answering

Unlike respondents that shirk by skipping items, some respondents provide answers, but they answer inconsistently. For instance, some respondents report the same answer to every question (i.e., straight-lining) in order to complete the survey with minimal effort and quickly (O’Conner, Sullivan, & Jones, 1982). Others simply provide random answers. Our second measure of survey effort aims to identify these patterns. That is, this measure aims to quantify the extent to which a respondent is carelessly submitting answers to surveys. Using this same type of measure, Hitt (2015) has shown that adolescents who engage in this type of behavior to larger degree have lower levels of educational attainment in adulthood.

We build this measure of careless answering by generalizing diagnostic techniques that psychologists have used to analyze data quality (Huang et al., 2012; Johnson, 2005; Meade & Craig, 2012). First we identify reliable self-reported scales that respondents had to answer. Again, we restrict ourselves to identifying scales in survey waves different from the waves that contain other data for our analysis to eliminate confounding variation. In this respect, we chose the following three scales to build our careless answering measure: A life satisfaction scale (included in wave survey 2), a well-being scale (included in wave survey 2) and a depression scale (included in wave survey 20). All these scales presented high reliability coefficients. In particular, the Cronbach's alphas for these scales were: 0.69 for life-satisfaction, 0.80 for well-being, and 0.87 for the depression scale. Answers among items in a reliable scale that are designed to capture a particular latent trait should be well correlated with each other. However, an individual who is careless in responding to a scale will submit answers that are not as well correlated with each other. Therefore in our second step, we regress responses from each item in a scale on the average score of the rest of items in the scale. Third, residuals from each of these regressions are obtained to capture the extent to which the response to a particular item is unpredictable, based upon the responses that the individual and others in the analytic sample provided for other items in the scale. Absolute values of each of these residuals are then standardized to account for any differences across the items within the same scale. These standardized residuals are then averaged within scales and standardized again to take into account differences across scales (e.g., different total number of items, answer options). Finally, a composite careless answering score is obtained by averaging these standardized averages residuals at the individual level⁶.

2.3 Outcome Measures, Cognitive Ability Measures and Other Relevant Information Available for our Analysis

Our outcome variables and other important controls were also constructed using multiple waves of the UAS (i.e., UAS1, UAS15). Our outcome variables include years of education imputed from the respondent's reported highest level of education achieved, as well as, highest level of education (i.e., less than high school, high school graduate, some college, college degree), household annual income in categories (less than \$25,000; \$25,000 to \$49,999; 50,000\$

⁶ See Hitt (2015) for additional technical details and explanation on this measure of careless answering.

to \$99,999; and more than \$100,000), a dummy that indicates that the respondent is currently employed, and a dummy for the respondent declaring to be occupied in a high-skilled job if employed. For building this last variable on occupational type we used information from the Bureau of Labor Statistics's Standard Occupational Classification system⁷ and information provided by Achieve through their study of the Future of the U.S Workforce⁸. With this information we classified reported job categories as high-skilled jobs (e.g., management, business and financial occupations or professional and related occupations) or not.

A very important control variable in our analysis is a measure of cognitive ability. This is so because we would like to study the correlation of our different measures of character skills and outcome variables above what would be driven by differences in cognitive ability. There are two sources of information for cognitive ability in the UAS that we use in this analysis. This include the 8 items Lipkus numeracy scale (Lipkus et al., 2001) and responses to a 5 items Cognitive reflection test (Frederick, S., 2005 and Toplak et al., 2014). The cognitive reflection test was developed with the aim to measure a specific cognitive ability. In particular, it measures respondent's ability to suppress and intuitive and spontaneous wrong answer in favor of a reflective and deliberative right response. We combined information in these two scales in a unique cognitive ability index using a factor analysis of the total number of correct responses in each of these tests. Both scales loaded into a unique factor with equal size weights.

Other important demographic information we use in our analysis include age, gender, ethnicity and 10 dummies for the region of residence.

3. Empirical Strategy for Validation of Measures

We take two approaches to validating our measures of character skills based on survey effort. First, we examine their partial correlation with self-reported measures, after controlling for cognitive ability and demographic information. Secondly, we study their correlation with education, income and labor outcomes and compare their predictive power with that of self-reported measures. The first approach provides an indication of construct validity while the second one provides an indication of criterion validity (Cronbach & Meehl, 1955).

⁷ http://www.bls.gov/soc/major_groups.htm

⁸ <http://www.achieve.org/future-us-workforce>

3.1 Construct Validity

Construct validity refers to the extent to which a measure of some latent trait captures what theory states it should capture. If the measure is correlated with other measures of the latent trait in question, then we say that the measure has convergent validity. Similarly, the measure should be uncorrelated with latent traits that have little to no relation with the latent trait in question. If so, the measure is said to have discriminant validity. We compute Spearman correlations for all of our measures and partial correlation coefficients, after controlling for cognitive ability and demographic information, to assess construct validity.

3.2 Criterion Validity

Measures of some latent trait ought to have some degree of criterion validity as well. That is, they should be correlated with other outcomes known to be correlated with the latent trait. In our case, we primarily have measures of character skills that we think are related to conscientiousness. If our measures do capture conscientiousness, then we should observe those measures to be related to outcomes such as educational attainment, earnings, and job status, as other research has demonstrated are linked to conscientiousness (Almlund et al. 2011).

For this analysis, we regress a variety of outcomes on each of the measures of character skills that are available in our data. The models also include controls for cognitive ability and demographic variables to disentangle their contribution to variation in the outcome variables. Linear regression models were used for continuous variables while multinomial or binary logit models were obtained for categorical dependent variables.

Admittedly, our sample consists of a cross-section of adults so we are only able to assess the contemporaneous correlation between measures of their character skills to our outcomes of interest. Ideally, researchers would administer measures of character skills to individuals prior to adulthood where long-run outcome variables are not determined yet and follow these individuals over time to determine the extent to which those measures are predictive of future outcomes, as done in Hitt, Trivitt, & Cheng (forthcoming) or Hitt (2015). Due to our sample, we are limited to exploring the concurrent validity, rather than the predictive validity, of our measures. Moreover, we cannot make causal claims about relationships that we find. It is possible that educational attainment, income, and other outcomes influence character skills through, for example, an increased opportunity cost of time. That being said, we believe that our sample provides a

valuable contribution to the research literature, given the breadth of available measures and the fact that it comprises a nationally-representative sample of adults. As discussed previously, validation of these measures of character skills have typically been done in samples of convenience, perhaps casting doubt on their generalizability. Moreover, no other dataset – cross-sectional or longitudinal – simultaneously has all the measures that are available in our data for similar validation purposes. We present the results of our analysis next.

4. Results

4.1 Summary Statistics

We first provide summary statistics for the sample of those who responded to our collected survey. Table 1 displays these statistics for outcome and demographic variables, using the sample of respondents to our designed survey module (UAS15), while Table 2 presents descriptive statistics for measures of character skills using any available data. All descriptive statistics and models in this paper are estimated using population weights to maintain representativeness of the U.S population. As we merged multiple survey modules to obtain information for our study, different models are estimated using as many observations as possible with sample sizes ranging from 974 to just over 1,700 respondents.

As can be seen in Table 1, the average age of our sample respondents is about 47 years old and slightly over half of the respondents are female. Most of our respondents are White while 15 and 13 percent of the respondents are Latino and African American, respectively. Only 4 percent are Asian and another 4 percent are of another ethnicity. Concerning levels of education, the majority of respondents have completed more than high school (58 percent). On the other hand, 31 percent of the respondents report having only completed high school while 11 percent have not completed high school. About 17 percent of respondents turned out to be in the highest level of household income. Approximately one third and one quarter of respondents report household incomes between \$50,000 and \$99,999 and between \$25,000 and \$49,000, respectively. Another quarter of respondents reported having household incomes lower than \$25,000. Finally, a majority of respondents (88 percent) report being employed with about 40 percent of them working in a high skilled job.

Similarly Table 2 shows summary statistics for the character-skill measures in our paper. Respondents to the UAS seem to present low item non-response rates of about 11 percent. It

should be pointed out that non-response is discouraged in the UAS. If respondents leave an answer blank they face a screen that reminds them how important their answers are and asks them to go back and provide a response. Obviously, respondents can choose to ignore the alert and continue answering subsequent items, hence the nonzero item nonresponse rates. Our careless answering measure, comprising a standardized average of the standardized residuals that capture inconsistent responses, ranges from -0.83 to 1.87 and presents a standard deviation of 0.47. This indicates considerable variation in the degree of care that respondents put in completing the surveys with some being more careful than the average (negative values) and some being considerably less careful (positive values).

The rest of variables in Table 2 correspond with self-reported measures of character skills that range from 1 to 5. Most of these variables present means in the range of 3 and 4. These values seem in line with the results reported by Duckworth & Quinn (2009) for a convenience sample of adults aged 25 and older.

4.2 Relationship among Character Skills Measures

Our first analysis studies the correlation between survey-based measures of character skills and the self-reported measures of personality, grit and growth mindset. Table 3 shows correlation coefficients across the different measures. Following Duckworth & Quinn (2009) we conducted a factor analysis on the responses to the grit scale and were able to replicate their results. In particular, we also found that two factors were identified from the scale corresponding to measures of passion and perseverance⁹. These predicted factors were also added to the list of variables to study correlations. Similarly, we coded different facets of conscientiousness, namely competence, orderliness, dutifulness, achievement striving, self-discipline and deliberateness (Jackson et al., 2010; MacCann, Duckworth, & Roberts, 2009). We studied correlations with each of these facets of conscientiousness separately.

We point out some of the patterns that are observed in the correlations presented in Table 3. As previously shown in the literature, self-reported grit appears to correlate most strongly with conscientiousness. However, it also presents moderate correlation with the personality factors neuroticism and agreeableness. Of the two identified factors within the grit scale, perseverance appears to be most strongly correlated with the other personality traits. Self-reported growth

⁹ Full results of the factor analysis are available from the authors upon request.

mindset presents some correlation with grit of the order of 0.1 but does not show significant correlation with any of the reported personality traits. Finally, concerning our survey effort measures of character skills, we observe that careless answering is correlated in the expected direction with self-reported grit ($\rho = -0.16$), indicating that those with higher levels of self-reported grit show lower levels of careless answering. Importantly, careless answering also presents a moderate correlation with conscientiousness ($\rho = -0.22$) and with neuroticism ($\rho = 0.33$), as self-reported grit did. This provides some evidence of the validity of our proposed careless answering measure because it seems to be capturing the expected personality traits. Measures of item non-response, however, do not seem to work as well. In this case, we fail to find much of a correlation between item non-response and grit, conscientiousness, or any of the other personality traits and character measures included.

In order to test if the above correlations are driven by cognitive ability or other demographic characteristics of the respondents, we also computed partial correlations that controlled for these variables. The results of this analysis is presented in Table 4. This partial correlation corroborated our previous results indicating that self-reported grit is positively correlated mostly with conscientiousness but also with neuroticism and agreeableness, while careless answering correlates most with neuroticism and conscientiousness. We again fail to find significant correlations between item non-response rates and any of the other character skill measures.

4.3 Relationship of Character Skills Measures and Education, Income and Labor Outcomes

Our second step in the validation analysis of survey based measures of character skills is to study their contemporaneous correlations with education, income and labor outcomes and to compare their predictive power with alternative self-reported measures.

Tables 5.A, 5.B and 5.C present the correlation between years of education and the different measures of character skills. Three estimates are presented in these tables: first without controls for cognitive ability, second with controls of cognitive ability, and finally with both controls of cognitive ability and demographic variables (age, gender, ethnicity and area of residence). As it can be seen in Table 5.A all the personality traits seem to correlate with final years of education. Table 5.B. shows the correlation with grit and growth mindset. As expected

we find a positive correlation between self-reported grit and years of education, with a standardized regression coefficient of about 0.25, even after controlling for cognitive ability and demographic information. From the two identified grit factors it seems that the factor related to perseverance is the one driving these results. Concerning growth mindset, however, we fail to find any significant correlation with years of education. Finally, Table 5.C presents the results for our measures based on survey effort. Careless answering presents a significant negative correlation with years of education and the estimated standardized coefficient is even higher than the one found for self-reported grit ($\beta = -0.40$). However, we do not find any correlation between item non-response rates and years of education.

We run an alternative model predicting educational attainment to account for the possibility of diploma effects. Tables 6.A-6.E present marginal effects for multinomial logit models for the highest level of education completed. Attainment levels are divided into four categories: less than high school, high school, some college, and college degree. Results generally comport with those found in Table 5. Interestingly, we find that self-reported grit – in particular, perseverance but not passion – is predictive of the probability of having a college degree (see Tables 6.C and 6.D). In contrast, careless answering patterns help us identify both those that do not have a high school diploma and those that finished college (see Table 6.E).

Turning to income and labor-market outcomes, Tables 7.A-7.E present marginal-effect estimates after running multinomial logit models that predict household income levels. These models control for cognitive ability, basic demographic information (age, gender, race and region of residence), highest level of education, and employment status. Cognitive ability presents a significant, negative and positive correlation with the probability of being in the low income and high income groups, respectively. Conscientiousness, however, does not seem to help explain household income above cognitive ability. The rest of personality factors, agreeableness, neuroticism, extroversion, openness do seem to explain household income but mostly through the probability of being in the highest level (Tables 7.A and 7.B). Growth mindset and grit do not seem to help us to significantly predict the level of household income, after controlling for cognitive ability (Tables 7.C and 7.D). Careless answering, however, presents a significant positive effect on the probability of having a low household income level (Table 7.E).

Tables 8.A-8.C present marginal effects for logit models estimating the probability of being employed, controlling for cognitive ability, basic demographic information (age, gender, race and region of residence), and highest level of education. As it can be seen in Table 8.A, although we find that some personality traits like neuroticism and agreeableness appear to be correlated with the probability of being employed, we fail to find a relationship between employment and conscientiousness. Similarly, we do not find a significant correlation between grit and the probability of employment. With careless answering, however, we find a significant negative effect for models without controls or that only control for cognitive ability but this effect disappears once demographic information and educational attainment is controlled for.

Finally, Tables 9.A-9.C present marginal effects for logit models for the probability of being employed in a high skilled occupation among those who reported being employed, controlling for cognitive ability, basic demographic information (age, gender, race and region of residence) and highest level of education. Concerning the Big 5 personality factors, we find that conscientiousness shows a significant correlation with the probability of having a high skilled job, among those employed, through its dutifulness and achievement facets. Self-reported grit also presents a positive and significant correlation with the probability of being employed in a high-skilled job. Both the passion and perseverance factors of self-reported grit are predictive of having a high skilled job. Essentially no relationship is found, however, for our measure of growth mindset. Careless answering, as self-reported grit, is significantly correlated with a lower probability of having a high skilled job. Finally, we also find a negative correlation between item non-response and the probability of having a high skilled job but this relationship is only marginally significant.

5. Further Discussion and Conclusions

There is now consensus in the literature that individual skills, like perseverance or self-control, play a prominent role on life-time outcomes like education and employment (Almlund et al., 2011). However, despite its potential importance, these skills are usually left unmeasured and not considered in policy evaluations. The main reason for this has been the difficulty of finding reliable ways to measure these skills. The few evaluations that have aimed to measure these important character skills have relied on self-reported measures. This type of measure, however, has been shown to suffer from important biases, like reference group bias and social desirability

bias. An alternative way to measure character skills would be through carefully designed behavioral tasks but these tasks result challenging and often expensive to perform in a large sample of respondents. This paper aims to validate and innovative way of measuring relevant character skills by studying measures of survey effort. We believe that a survey can be seen as a behavioral task and that by observing the effort that individuals put forward on answering we can gather relevant information about respondent's character skills.¹⁰

We make use of the Understanding America Study, a unique nationally representative internet panel, to study the performance of survey based effort measures of cognitive skills. We study their correlation with self-reported measures of similar skills, as well as education, income and labor outcomes. In particular, we study two survey effort measures these being: item non-response rates and measures of careless answering.

Our results show that careless answering measures show promise to be a valid proxy of relevant character skills. In particular, they exhibit correlations with self-reported measures of grit, conscientiousness and neuroticism. Similarly, self-reported grit also presents correlations with conscientiousness and neuroticism, but also appears correlated with agreeableness. Just as self-reported grit, careless answering is correlated with educational attainment. In an analysis of levels of education, however, we observe that while self-reported grit helps us predict the probability of having a college degree, careless answering is useful to predict both the probability of low levels of education as well as college. Similarly, while self-reported grit does not present a correlation with household income levels, after controlling for educational attainment and employment status, careless answering helps us predict the probability of having low household income. A weaker relationship is observed between careless answering and the probability of being employed, after controlling for cognitive ability and educational levels. This result goes in line with prior research suggesting that the effect of character skills on labor-market outcomes attenuates after controlling for educational attainment. That is, benefits to labor-market outcomes that are due to having higher levels of particular character skills accrue through the effects of character skills on educational attainment (Cawley, Heckman & Vytlačil, 2001 or Mendez & Zamarro, 2015). Finally, both self-reported grit and careless answering show

¹⁰ We also wish to point out that our survey-effort measures can be used to analyze existing data sets that did not collect measures of character skills. Measures of character skills can be recovered in in these data sets, opening up the possibility of studying character skills.

significant correlations with the probability of having a high skilled job, among those employed, after controlling for educational attainment, demographic information and cognitive ability. That the predictive power of the survey-effort measures persists even after controlling for educational attainment contrasts with that of self-reported measures.

Item non-response does not seem to show any correlation with the self-reported measures of character skills considered in this paper, or with education and labor outcomes. This is in contrast with results by Hitt, Trivitt and Cheng (forthcoming) who found that when measured among school-aged children, item non-response rates in paper-pencil surveys carried out in schools were predictive of later life education and employment outcomes. One possible explanation for our different result is the fact that our data comes from a survey that purposely tried to minimize non-response, reducing variation in item nonresponse. Also, our survey was performed on adults over the internet and this result might indicate either the effects of survey mode, or the fact that this was applied to a different population in a very different context.

Although more research is needed to understand the full potential of survey effort measures of character skills, this paper complements previous papers (Borghans & Schils, 2015; Hitt, 2015; Hitt, Trivitt & Cheng, forthcoming; Mendez, Zamarro, Clavel & Hitt, 2015) that found measures of survey effort in adolescents to be predictive of short- and long-run life outcomes. To our knowledge, this is also the first time that some of the self-reported measures of character skills are studied in a nationally representative sample, as opposed to multiple convenient samples. We suggest that our results show that careless answering measures are a reasonable proxy for a dimension of character skills related to consciousness and neuroticism which explains variability in education, income and labor outcomes above and beyond what is explained by cognitive ability and demographic information.

In conclusion, we return to the main motivation behind validating survey-effort measures of character skills. Complete evaluations of policy or program interventions require researchers to consider their effects not only on student performance on tests of cognitive ability but also on student character skills. Ignoring the latter would have understated the benefits that various policies and programs bequeathed to students and other stakeholders. Scholars are only recently beginning to recognize this fact but nonetheless face difficulty in obtaining viable measures of character skills for research purposes. The study and development of survey-effort measures is

intended to address this data limitation. We welcome additional research into improving measures of character skills in the same spirit as this analysis.

References

- Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). *Personality psychology and economics* (No. w16822). National Bureau of Economic Research.
- Becker, Gary. Human capital; a theoretical and empirical analysis, with special reference to education. New York: Columbia University Press, 1964.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development*, 78(1), 246-263.
- Borghans, Lex, and Trudie Schils. 2015. "The Leaning Tower of Pisa." Working Paper. Accessed February 24. <http://www.sole-jole.org/13260.pdf>
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4), 972-1059.
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, 343(6178), 1478-1485.
- Cawley, J., Heckman, J. J., & Vytlacil, E.J. (2001). Three observations on wages and measured cognitive ability. *Labour Economics* 8 (4), 419-442.
- Chingos, M. M., & Peterson, P. E. (2015). Experimentally estimated impacts of school vouchers on college enrollment and degree attainment. *Journal of Public Economics*, 122, 1-12.
- Cowen, J. M., Fleming, D. J., Witte, J. F., Wolf, P. J., & Kisida, B. (2013). School vouchers and student attainment: Evidence from a state-mandated study of Milwaukee's parental choice program. *Policy Studies Journal*, 41(1), 147-168.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- Dobbie, W., & Fryer, R. G. (2015). The Medium-term Impacts of High-achieving Charter Schools. *Journal of Political Economy*, 123(5), 985–1037.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 92(6), 1087.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of personality assessment*, 91(2), 166-174.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- Falk, A., & Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535-538.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25-42.
- Funder, D. (2001). Personality. *Annual Review of Psychology*, 52(1), 197-221.
- Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125-149.
- Grant, H. & Dweck, C.S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85, 541-553.
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267-321.

- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior*. *Journal of Labor Economics*, 24(3), 411-482.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative economics*, 1(1), 1-46.
- Heckman, J. J., Pinto, R., & Savelyev, P. A. (2013). *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes*. *American Economic Review*, 103(6), 2052-86.
- Hitt, C. E., Cheng, A., & Julie, R. T. (forthcoming). When You Say Nothing at All: The Surprising Predictive Power of Student Effort on Surveys. *Economics of Education Review*.
- Huang, J.L., Curran, P.G., Keeney, J., Paposki, E.M., DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114.
- Ingels, S. J., Pratt, D. J., Rogers, J. E., Siegel, P. H., & Stutts, E. S. (2004). Education Longitudinal Study of 2002: Base Year Data File User's Manual. NCES 2004-405. *National Center for Education Statistics*.
- Jackson, K. (2012). *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina* (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.
- Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the behavioral indicators of conscientiousness (BIC). *Journal of Research in Personality*, 44(4), 501–511.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley: *University of California, Berkeley*, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3, 114-158.
- John, O.P. (1990). The "Big Five factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L.A. Pervin (ed.), *Handbook of personality theory and research* (pp. 66-100). New York : Guilford.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999), 102-138.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 39, 103–129.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1), 37-44.
- MacCann, C., Duckworth, A.L., & Roberts, R.D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, 19(4), 451–458.
- Meade, A., & Bartholomew, C.S. (2012). Identifying careless responses survey data. *Psychological Methods*, 17(3), 437-455.
- Mendez, I. & Zamarro, G. (2015). The intergenerational transmission of noncognitive skills and their effect on education and employment outcomes. EDRE Working Paper 2015-04, Fayetteville, AR: Department of Education Reform, University of Arkansas.

- Mendez, I., Zamorro, G., Clavel, J. G. & Hitt, C. (2015). Non-cognitive Abilities and Spanish Regional Differences in Student Performance in PISA 2009. EDRE Working Paper 2015-05, Fayetteville, AR: Department of Education Reform, University of Arkansas.
- Mischel, W., & Ebbesen, E.B. (1970). Attention in Delay of Gratification. *Journal of Personality and Social Psychology*, 16(2), 329-337.
- O'Connor, P. J., Sullivan, G. L., & Jones, W. H. (1982). An Evaluation Of The Characteristics Of Response Quality Induced By Follow-Up Survey Methods. *Advances in Consumer Research*, 9(1).
- Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, 64(2), 169-190.
- Peters, E. (2008). Numeracy and the perception and communication of risk. *Annals of the New York Academy of Sciences*, 1128(1), 1-7.
- Sass, T., Zimmer, R., Gill, B., & Booker, K. (in press). Charter high schools' effect on long-term attainment and earnings. *Journal of Policy Analysis and Management*.
- Weel, B. ter. (2008). The Noncognitive Determinants of Labor Market and Behavioral Outcomes: Introduction to the Symposium. *Journal of Human Resources*, 43(4), 729–737.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147-168.
- West, M., Kraft, M. A., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148–170.
- Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo., L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management*, 32 (2), 246-270.

Table 1. Summary Statistics for Demographic and Outcome Variables

Measure	Observations	Mean	Standard Deviation	Minimum	Maximum
Age	1692	47.0	16.50	18	100
Female	1702	0.51	0.50	0	1
<i>Racial Background</i>					
Hispanic/Latino	1698	0.15	0.36	0	1
Black	1702	0.13	0.34	0	1
Asian	1703	0.04	0.20	0	1
Native American	1702	0.04	0.19	0	1
Pacific Islander	1702	0.00	0.05	0	1
<i>Educational Attainment</i>					
College Graduate	1702	0.30	0.46	0	1
Some College	1702	0.28	0.45	0	1
High School	1702	0.31	0.46	0	1
Less than High School	1684	0.11	0.31	0	1
<i>Income Levels</i>					
More than \$100,000	1684	0.17	0.38	0	1
\$50,000 to \$99,999	1684	0.34	0.47	0	1
\$25,000 to \$49,999	1684	0.24	0.43	0	1
Less than \$25,000	1684	0.26	0.44	0	1
<i>Employment Status</i>					
Employed	1219	0.88	0.33	0	1
Employed at a High Skilled Job	974	0.40	0.49	0	1

Note: Summary statistics presented using population weights.

Table 2. Summary Statistics for Measures of Character Skills

Measure	Observations	Mean	Standard Deviation	Minimum	Maximum
Item non-response	1433	0.11	0.03	0.01	0.32
Careless Answers	1703	0.01	1.02	-1.95	4.39
Grit	1705	3.60	0.57	1.38	5.00
Growth Mindset	1695	3.67	0.99	1.00	5.00
Agreeableness	1698	4.03	0.59	1.67	5.00
Neuroticism	1698	2.64	0.81	1.00	5.00
Openness	1696	3.60	0.63	1.60	5.00
Extroversion	1697	3.36	0.79	1.00	5.00
Conscientiousness	1699	4.06	0.61	1.00	5.00
Competence	1696	3.97	0.85	1.00	5.00
Order	1693	4.17	0.92	1.00	5.00
Dutifulness	1695	4.21	0.70	1.00	5.00
Achievement Striving	1698	4.33	0.76	1.00	5.00
Self-discipline	1681	4.00	1.16	1.00	5.00
Deliberation	1695	3.35	1.24	1.00	5.00

Note: Summary statistics presented using population weights.

Table 3. Correlations across Alternative Character Skills Measures

	Grit	Passion	Perseverance	Growth Mindset	Conscientiousness	Competence	Orderliness	Dutifulness	Achievement striving	Self-discipline	Deliberateness	Agreeableness	Neuroticism	Extroversion	Openness	Item non-response Rate
Growth Mindset	0.10	0.10	0.05	-												
Conscientiousness	0.48	0.26	0.47	0.05	-											
Competence	0.38	0.21	0.38	0.00	-	-										
Orderliness	0.29	0.14	0.29	0.00	-	0.45	-									
Dutifulness	0.28	0.14	0.29	0.03	-	0.44	0.32	-								
Achievement Striving	0.33	0.14	0.40	0.04	-	0.43	0.40	0.31	-							
Self-discipline	0.33	0.17	0.33	0.07	-	0.42	0.29	0.36	0.30	-						
Deliberateness	0.35	0.28	0.21	0.09	-	0.41	0.28	0.36	0.22	0.38	-					
Agreeableness	0.27	0.14	0.25	0.09	0.44	0.27	0.36	0.39	0.30	0.27	0.28	-				
Neuroticism	-0.33	-0.4	-0.21	-0.09	-0.43	-0.29	-0.24	-0.35	-0.18	-0.30	-0.41	-0.39	-			
Extroversion	0.18	0.08	0.16	0.00	0.29	0.22	0.20	0.09	0.28	0.25	0.14	0.19	-0.25	-		
Openness	0.14	0.05	0.16	0.03	0.27	0.18	0.28	0.14	0.33	0.09	0.10	0.25	-0.19	0.33	-	
Item non-response Rate	0.01	-0.01	0.04	0.01	-0.01	0.02	-0.04	-0.03	-0.02	0.01	0.00	-0.05	0.05	-0.02	-0.02	-
Careless Answers	-0.16	-0.11	-0.12	-0.02	-0.22	-0.16	-0.17	-0.19	-0.15	-0.11	-0.14	-0.12	0.32	-0.09	-0.12	0.02

Note: N= 1,412. Correlations presented using population weights.

Table 4. Partial Correlations across Character Skills Measures

	Grit	Item non-response rate	Careless Answers
Grit	-	0.00	-0.15
Item non-response rate	0.00	-	0.02
Careless Answers	-0.15	0.02	-
Passion	-	0.01	-0.07
Perseverance	-	-0.01	-0.15
Growth Mindset	0.10	0.02	-0.02
Conscientiousness	0.46	-0.01	-0.21
Competence	0.36	0.01	-0.16
Orderliness	0.31	-0.03	-0.16
Dutifulness	0.28	0.03	-0.18
Achievement Striving	0.34	-0.04	-0.13
Self-discipline	0.32	0.00	-0.10
Deliberateness	0.33	0.01	-0.12
Agreeableness	0.26	-0.03	-0.15
Neuroticism	-0.33	0.03	0.28
Extroversion	0.17	-0.01	-0.08
Openness	0.15	-0.03	-0.05

Note: Partial correlations presented using population weights. Controls for cognitive ability & demographic variables included.

Table 5.A. Years of Education and the Big- 5 Personality Factors

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		1.203** (0.074)	1.181** (0.076)		1.175** (0.073)	1.175** (0.075)
Conscientiousness	0.191* (0.095)	0.185* (0.081)	0.194* (0.078)			
Agreeableness	-0.279** (0.095)	-0.005 (0.093)	-0.059 (0.095)	-0.310** (0.097)	-0.027 (0.093)	-0.062 (0.094)
Neuroticism	-0.253** (0.088)	-0.040 (0.077)	-0.158* (0.081)	-0.315** (0.089)	-0.077 (0.079)	-0.170** (0.080)
Extroversion	-0.326** (0.092)	-0.158† (0.084)	-0.184* (0.078)	-0.295** (0.088)	-0.129 (0.080)	-0.156** (0.077)
Openness	0.389** (0.077)	0.222** (0.073)	0.263** (0.073)	0.332** (0.081)	0.192** (0.076)	0.248** (0.076)
Competence				0.031 (0.103)	0.065 (0.088)	0.067 (0.087)
Order				-0.011 (0.096)	-0.018 (0.086)	-0.085 (0.079)
Dutifulness				0.265* (0.105)	0.173† (0.096)	0.119 (0.091)
Achievement Striving				0.329** (0.088)	0.227** (0.075)	0.228** (0.072)
Discipline				-0.191* (0.090)	-0.159† (0.085)	-0.103 (0.077)
Deliberation				-0.165 (0.104)	-0.035 (0.093)	0.047 (0.089)
Demographic Variables Included			x			x
Observations	1,695	1,695	1,654	1,687	1,687	1,647
Adjusted R ²	0.038	0.224	0.270	0.058	0.233	0.277

Note: Standardized regression coefficients reported. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 5.B. Years of Education, Grit & Growth Mindset

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cognitive Ability		1.241** (0.073)	1.236** (0.076)		1.247** (0.074)	1.241** (0.076)		1.237** (0.074)	1.231** (0.077)
Grit	0.201* (0.095)	0.211* (0.084)	0.254** (0.080)						
Passion				0.070 (0.096)	0.087 (0.083)	0.143† (0.080)			
Perseverance				0.247** (0.074)	0.242** (0.068)	0.233** (0.066)			
Growth Mindset							-0.125 (0.084)	-0.124 (0.076)	-0.112 (0.074)
Demographic Variables Included			x			x			x
Observations	1,701	1,701	1,662	1,697	1,697	1,658	1,691	1,697	1,658
Adjusted R ²	0.005	0.218	0.263	0.009	0.224	0.266	0.002	0.214	0.255

Note: Standardized regression coefficients reported. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 5.C. Years of Education & Survey Effort Measures of Character Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		1.325** (0.080)	1.340** (0.083)		1.181** (0.076)	1.168** (0.075)
Nonresponse	-0.038 (0.126)	-0.033 (0.101)	0.003 (0.104)			
Careless Answering				-0.496** (0.074)	-0.278** (0.075)	-0.386** (0.071)
Demographic Variables Included			x			x
Observations		1,435	1,435	1,396	1,702	1,661
Adjusted R ²		-0.001	0.237	0.280	0.038	0.275

Note: Standardized regression coefficients reported. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 6.A. Educational Level and the Big-5 Personality Factors

	Low	Medium	Medium-high	High
Cognitive Ability	-0.074** (0.020)	-0.125** (0.020)	0.022 (0.015)	0.177** (0.011)
Conscientiousness	-0.025* (0.013)	0.023 (0.018)	-0.023 (0.015)	0.025† (0.013)
Agreeableness	-0.008 (0.014)	0.010 (0.020)	0.008 (0.016)	-0.010 (0.015)
Neuroticism	0.019 (0.012)	-0.004 (0.018)	0.002 (0.016)	-0.017 (0.013)
Extroversion	0.024† (0.012)	-0.010 (0.018)	0.002 (0.014)	-0.017 (0.012)
Openness	-0.021† (0.012)	-0.032† (0.017)	0.025† (0.014)	0.027* (0.013)

Note: N = 1,655. Pseudo R² = 0.131. Table reports average marginal effects estimated after running multinomial logit models. Demographic variables included in controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 6.B. Education Level and the Big- 5 personality Factors

	Low	Medium	Medium-high	High
Cognitive Ability	-0.074** (0.019)	-0.119** (0.020)	0.018 (0.015)	0.176** (0.011)
Competence	-0.023 (0.014)	0.008 (0.021)	0.001 (0.017)	0.014 (0.014)
Order	0.006 (0.011)	0.026 (0.019)	-0.028† (0.015)	-0.005 (0.013)
Dutifulness	-0.012 (0.012)	0.001 (0.020)	0.019 (0.018)	-0.008 (0.016)
Achievement Striving	-0.008 (0.010)	-0.039* (0.018)	0.017 (0.017)	0.030* (0.014)
Discipline	0.025† (0.013)	0.009 (0.019)	-0.035* (0.016)	0.001 (0.013)
Deliberation	-0.024† (0.013)	0.035† (0.019)	-0.010 (0.016)	-0.000 (0.013)
Agreeableness	-0.010 (0.014)	0.012 (0.020)	0.007 (0.016)	-0.008 (0.015)
Neuroticism	0.011 (0.012)	0.008 (0.019)	0.001 (0.016)	-0.020 (0.013)
Extroversion	0.018 (0.012)	-0.006 (0.018)	0.007 (0.015)	-0.018 (0.012)
Openness	-0.022† (0.012)	-0.025 (0.017)	0.023 (0.014)	0.024† (0.013)

Note: N=1,648. Pseudo $R^2 = 0.136$. Table reports average marginal effects estimated after running multinomial logit models. Demographic variables included in controls. Estimates use population weights. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

Table 6.C. Educational level and Grit

	Less than High School	High School	Some College	College	Less than High School	High School	Some College	College
Cognitive Ability	-0.078** (0.020)	-0.125** (0.019)	0.020 (0.015)	0.183** (0.011)	-0.080*** (0.020)	-0.124*** (0.019)	0.021 (0.015)	0.183*** (0.011)
Grit	-0.016 (0.011)	-0.016 (0.016)	-0.002 (0.013)	0.033** (0.011)				
Passion					-0.004 (0.010)	-0.001 (0.016)	-0.009 (0.013)	0.015 (0.011)
Perseverance					-0.022** (0.009)	-0.020 (0.014)	0.009 (0.013)	0.033*** (0.011)
Observations		1,663					1,659	
Pseudo R ²		0.141					0.146	

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables included in controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 6.D. Educational level and Growth Mindset

	Less than High School	High School	Some College	College
Cognitive Ability	-0.077** (0.020)	-0.125** (0.019)	0.020 (0.015)	0.182** (0.011)
Growth Mindset	0.012 (0.012)	-0.006 (0.016)	0.006 (0.014)	-0.012 (0.011)
Observations		1,695		
Pseudo R ²		0.138		

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables included in controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 6.E. Educational level and Survey Effort Measures of Character Skills

	Less than High School	High School	Some College	College	Less than High School	High School	Some College	College
Cognitive Ability	-0.087** (0.022)	-0.123** (0.021)	0.019 (0.016)	0.191** (0.011)	-0.065** (0.019)	-0.127** (0.019)	0.020 (0.015)	0.172** (0.010)
Nonresponse	0.023 (0.015)	-0.038† (0.021)	0.004 (0.019)	0.010 (0.015)				
Careless Answering					0.048** (0.011)	0.006 (0.016)	0.018 (0.013)	-0.072** (0.012)
Observations		1,396				1,662		
Pseudo R ²		0.153				0.157		

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables included in controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 7.A. Household Income Level and the Big- 5 personality Factors

	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000
Cognitive Ability	-0.049** (0.015)	-0.009 (0.019)	-0.012 (0.020)	0.069** (0.013)
Conscientiousness	-0.015 (0.017)	-0.023 (0.019)	0.025 (0.023)	0.012 (0.014)
Agreeableness	-0.007 (0.015)	0.025 (0.018)	0.017 (0.020)	-0.035** (0.012)
Neuroticism	0.007 (0.015)	0.012 (0.019)	0.014 (0.021)	-0.033* (0.014)
Extroversion	-0.011 (0.013)	0.007 (0.018)	-0.028 (0.019)	0.031* (0.014)
Openness	0.022† (0.013)	0.042* (0.017)	-0.023 (0.018)	-0.041** (0.014)

Note: N = 1,172. Pseudo R² = 0.220. Table reports average marginal effects estimated after running multinomial logit models. Demographic variables, educational attainment levels, and employment status included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 7.B. Household Income Level and the Big- 5 personality Factors

	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000
Cognitive Ability	-0.048** (0.015)	-0.009 (0.019)	-0.012 (0.020)	0.070** (0.013)
Competence	-0.003 (0.015)	-0.015 (0.020)	0.018 (0.021)	0.001 (0.017)
Order	0.005 (0.012)	0.034† (0.017)	-0.048* (0.020)	0.010 (0.014)
Dutifulness	0.004 (0.017)	-0.018 (0.020)	-0.002 (0.023)	0.015 (0.017)
Achievement Striving	-0.018 (0.015)	-0.006 (0.018)	0.057* (0.024)	-0.033* (0.015)
Discipline	-0.017 (0.016)	-0.024 (0.018)	0.019 (0.021)	0.022 (0.014)
Deliberation	0.009 (0.016)	0.000 (0.019)	-0.010 (0.020)	0.001 (0.014)
Agreeableness	-0.008 (0.015)	0.023 (0.018)	0.021 (0.021)	-0.036** (0.013)
Neuroticism	0.013 (0.016)	0.008 (0.019)	0.008 (0.020)	-0.030* (0.014)
Extroversion	-0.004 (0.014)	0.009 (0.018)	-0.035† (0.020)	0.031* (0.014)
Openness	0.021† (0.013)	0.034* (0.017)	-0.020 (0.018)	-0.036** (0.013)

Note: N= 1,168. Pseudo R² = 0.153. Table reports average marginal effects estimated after running multinomial logit models. Demographic variables, educational attainment levels, and employment status included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 7.C. Household Income and Grit

	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000
Cognitive Ability	-0.046** (0.015)	-0.018 (0.019)	-0.007 (0.020)	0.071** (0.014)				
Grit	0.008 (0.013)	-0.022 (0.016)	-0.000 (0.019)	0.014 (0.015)				
Passion					0.013 (0.011)	-0.026 (0.017)	0.011 (0.019)	0.001 (0.014)
Perseverance					-0.013 (0.013)	0.000 (0.016)	-0.007 (0.019)	0.021 (0.013)
Observations		1,181				1,178		
Pseudo R ²		0.192				0.208		

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables, educational attainment levels, and employment status included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 7.D. Household Income and Growth Mindset

	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000
Cognitive Ability	-0.045** (0.015)	-0.018 (0.019)	-0.006 (0.020)	0.069** (0.013)
Grit	0.014 (0.013)	-0.018 (0.015)	-0.003 (0.017)	0.007 (0.012)
Growth Mindset				
Observations		1,175		
Pseudo R ²		0.191		

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables, educational attainment levels, and employment status included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 7.E. Household Income and Survey Effort Measures of Character Skills

	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000	Less than \$25,000	\$25,000 to \$49,999	\$50,000 to \$99,999	More than \$100,000
Cognitive Ability	-0.050** (0.016)	-0.031 (0.020)	-0.002 (0.021)	0.083** (0.014)	-0.045** (0.014)	-0.017 (0.019)	-0.009 (0.019)	0.070** (0.014)
Nonresponse	-0.040* (0.020)	0.020 (0.018)	-0.010 (0.028)	0.030† (0.016)				
Careless Answering					0.054** (0.013)	-0.008 (0.019)	-0.034 (0.021)	-0.027 (0.017)
Observations		986				1,180		
Pseudo R ²		0.220				0.202		

Note: Table reports average marginal effects estimated after running multinomial logit models. Demographic variables, educational attainment levels, and employment status included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 8. A. Employment Status & the Big- 5 personality Factors

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		0.041** (0.015)	0.004 (0.015)		0.038** (0.015)	0.004 (0.015)
Conscientiousness	0.038 (0.023)	0.040† (0.024)	0.024 (0.023)			
Agreeableness	-0.055* (0.025)	-0.042† (0.025)	-0.037† (0.022)	-0.062* (0.026)	-0.049† (0.026)	-0.043* (0.021)
Neuroticism	-0.047* (0.018)	-0.039* (0.019)	-0.036† (0.018)	-0.056** (0.019)	-0.048* (0.020)	-0.041* (0.020)
Extroversion	0.009 (0.017)	0.015 (0.018)	0.022 (0.017)	0.007 (0.018)	0.012 (0.019)	0.020 (0.018)
Openness	-0.033 (0.021)	-0.041† (0.021)	-0.047* (0.020)	-0.041* (0.021)	-0.048* (0.022)	-0.051* (0.021)
Facet: Competence				-0.015 (0.020)	-0.015 (0.020)	-0.009 (0.017)
Facet: Order				0.011 (0.014)	0.010 (0.014)	0.018 (0.014)
Facet: Dutifulness				0.030 (0.022)	0.028 (0.021)	0.027 (0.019)
Facet: Achievement				0.041* (0.020)	0.038† (0.020)	0.012 (0.017)
Facet: Discipline				0.002 (0.013)	0.005 (0.013)	-0.002 (0.011)
Facet: Deliberation				-0.011 (0.013)	-0.010 (0.013)	-0.013 (0.015)
Demographic Variables						
Included			x			x
Observations	1,214	1,214	1,172	1,210	1,210	1,168
Pseudo R ²	0.031	0.047	0.160	0.044	0.058	0.166

Note: Table reports average marginal effects estimated after running logit models. Demographic variables and educational attainment levels are included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 8.B. Employment Status, Grit & Growth Mindset

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cognitive Ability		0.047** (0.014)	0.008 (0.014)		0.047** (0.014)	0.010 (0.014)		0.048** (0.015)	0.008 (0.014)
Grit	0.018 (0.013)	0.017 (0.013)	0.017 (0.011)						
Passion				-0.003 (0.013)	-0.003 (0.013)	-0.001 (0.012)			
Perseverance				0.032* (0.013)	0.032* (0.014)	0.031* (0.014)			
Growth Mindset							0.010 (0.013)	0.010 (0.013)	0.018 (0.012)
Demographic Variables Included			x			x			x
Observations	1,221	1,221	1,181	1,218	1,218	1,175	1,215	1,215	1,175
Pseudo-R ²	0.004	0.027	0.137	0.012	0.036	0.144	0.002	0.026	0.140

Note: Table reports average marginal effects estimated after running logit models. Demographic variables and educational attainment levels are included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 8.C. Employment Status & Survey Effort Measures of Character Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		0.043** (0.016)	0.014 (0.016)		0.034* (0.015)	0.006 (0.014)
Nonresponse	-0.022 (0.018)	-0.023 (0.018)	-0.001 (0.017)			
Careless Answering				-0.047** (0.013)	-0.042** (0.013)	-0.018 (0.014)
Demographic Variables Included			x			x
Observations	1,024	1,024	984	1,222	1,222	1,180
Pseudo-R ²	0.004	0.024	0.127	0.027	0.040	0.139

Note: Table reports average marginal effects estimated after running logit models. Demographic variables and educational attainment levels are included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 9. A. High Skilled Occupation & the Big- 5 personality Factors

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		0.154** (0.018)	0.073** (0.020)		0.153** (0.017)	0.071** (0.020)
Conscientiousness	0.005 (0.045)	0.022 (0.042)	0.043 (0.037)			
Agreeableness	-0.060 (0.043)	-0.003 (0.042)	-0.022 (0.036)	-0.085* (0.041)	-0.023 (0.041)	-0.035 (0.035)
Neuroticism	-0.051 (0.034)	-0.014 (0.032)	-0.038 (0.028)	-0.079* (0.032)	-0.037 (0.031)	-0.050† (0.027)
Extroversion	-0.025 (0.032)	0.013 (0.032)	0.008 (0.027)	-0.024 (0.032)	0.018 (0.033)	0.011 (0.028)
Openness	0.099** (0.037)	0.061† (0.037)	0.036 (0.033)	0.073* (0.037)	0.040 (0.037)	0.022 (0.034)
Facet: Competence				-0.010 (0.033)	-0.003 (0.029)	-0.019 (0.026)
Facet: Order				0.023 (0.029)	0.022 (0.026)	0.024 (0.023)
Facet: Dutifulness				0.071* (0.036)	0.070* (0.036)	0.066* (0.029)
Facet: Achievement				0.083** (0.038)	0.077** (0.035)	0.066* (0.033)
Facet: Discipline				-0.035 (0.022)	-0.044† (0.021)	-0.021 (0.018)
Facet: Deliberation				-0.056** (0.020)	-0.037* (0.019)	-0.022 (0.017)
Demographic Variables Included			x			x
Observations	853	853	830	851	851	828
Pseudo-R ²	0.015	0.092	0.251	0.040	0.114	0.264

Note: Table reports average marginal effects estimated after running logit models. Analytic sample is restricted to those who are employed. Demographic variables and educational attainment levels are included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 9. B. High Skilled Occupation, Grit & Growth Mindset

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cognitive Ability		0.161** (0.016)	0.087** (0.019)		0.159** (0.016)	0.088** (0.019)		0.156** (0.017)	0.080** (0.020)
Grit	0.080** (0.021)	0.088** (0.020)	0.077** (0.017)						
Passion				0.047* (0.022)	0.050* (0.020)	0.049** (0.017)			
Perseverance				0.064** (0.023)	0.074** (0.021)	0.059** (0.018)			
Growth Mindset							-0.032 (0.022)	-0.032 (0.021)	-0.028 (0.018)
Demographic Variables Included			x			x			x
Observations	860	860	837	858	858	835	856	856	833
Pseudo-R ²	0.018	0.112	0.266	0.017	0.110	0.272	0.003	0.090	0.246

Note: Table reports average marginal effects estimated after running logit models. Analytic sample is restricted to those who are employed. Demographic variables and educational attainment levels are included as controls. Estimates use population weights. †p<0.1; *p<0.05; **p<0.01.

Table 9.C. High Skilled Occupation & Survey Effort Measures of Character Skills

	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Ability		0.160** (0.017)	0.065** (0.021)		0.147** (0.017)	0.078** (0.020)
Nonresponse	-0.057 (0.035)	-0.059† (0.034)	-0.046 (0.030)			
Careless Answering				-0.086** (0.023)	-0.055* (0.023)	-0.056** (0.020)
Demographic Variables Included			x			x
Observations	754	754	731	860	860	837
Pseudo-R ²	0.005	0.098	0.270	0.018	0.095	0.252

Note: Table reports average marginal effects estimated after running logit models. Analytic sample is restricted to those who are employed. Demographic variables and educational attainment levels are included as controls. Estimates use population weights.

†p<0.1; *p<0.05; **p<0.01.