

# The Softer Side of Learning: Measuring Students' Non-Cognitive Skills

Anna J. Egalite<sup>1</sup>  
Harvard Kennedy School

Jonathan N. Mills  
University of Arkansas

Jay P. Greene  
University of Arkansas

*EDRE Working Paper No. 2014-03*

*Last Updated July 2014*

*Please Do Not Cite Without Author Permission*

## Abstract

With an abundance of datasets of standardized test score data, researchers and education policymakers run the risk of focusing exclusively on the measurement of cognitive outcomes in key academic subjects such as math and reading at the expense of important non-cognitive outcomes. We use behavioral measures of conscientiousness, perseverance, and delay of gratification as well as a self-reported measure of student grit— defined as student perseverance and passion for long-term goals— to assess the non-cognitive skills of 174 11<sup>th</sup> and 12<sup>th</sup> grade students attending a residential public high school in Arkansas. Analysis shows that 11<sup>th</sup> grade students rate themselves lower on self-reported grit but outperform their 12<sup>th</sup> grade counterparts on three behavioral measures of persistence, delay of gratification, and conscientiousness. These findings point to the strengths and limitations of existing tools for measuring non-cognitive skills and the need for more measures to be developed and tested with diverse populations.

*Keywords:* Non-cognitive skills; conscientiousness; perseverance; grit

---

<sup>1</sup> Corresponding author. Email: [Anna\\_Egalite@hks.harvard.edu](mailto:Anna_Egalite@hks.harvard.edu)

## The Softer Side of Learning: Measuring Students' Non-Cognitive Skills

Growing research, accompanied by a surge in popular interest (Tough, 2012; Whitman, 2008), has demonstrated the importance of non-cognitive skills for student outcomes (De Ridder, Lensvelt-Mulders, Finkenauer, Stok & Baumeister, 2012; Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Seligman, 2005; Duckworth, Tsukayama, & May, 2010; Peterson & Seligman, 2004). The term “non-cognitive skills” refers to personality traits that have been shown to be associated with individuals’ success. It incorporates constructs such as optimism, resilience, adaptability, and conscientiousness. The move towards a larger emphasis on non-cognitive skills has been fueled by research demonstrating that these types of skills are responsive to educational intervention (Cunha & Heckman, 2008; Dee & West, 2011; Jackson, 2012) and are related to important future outcomes such as effective social skills (Mischel, Shoda, & Peake, 1988), improved performance on academic tasks (Blackwell, Trzaesniewski, & Dweck, 2007; Mischel, Shoda & Rodriguez, 1989; Wulfert et al, 2002; Zimmerman, 2000), reduced involvement with cigarettes, alcohol, and marijuana (Wulfert, Block, Santa Ana, Rodriguez, & Colman, 2002), better physical health (Moffitt, Arseneault, Belsky, Dickson, Hancox, Harrington, Houts, Poulton, Roberts, Ross, Sears, Thomson and Caspi, 2011) and improved ability to handle challenging situations in adulthood (Mischel, Shoda, & Peake, 1988; Mischel, Shoda, & Peake, 1989; Shoda, Mischel, & Peake, 1990).

Across the United States, educators and policymakers involved with the K-16 pipeline are recognizing the importance of traits such as persistence, resilience, adaptability, and grit as complements to traditional measures of academic achievement assessed via standardized tests. From preschool to college, school leaders are “sweating the small stuff” (Whitman, 2008) and

implementing programs to foster students' non-cognitive traits (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

Non-cognitive skills have traditionally been a focus for preschool and kindergarten programs (Daily, Burkhauser, and Halle, 2010; Goldstein, Eastwood, and Behuniak, 2014). Head Start CARES (Classroom-based Approaches and Resources for Emotion and Social skill promotion) is just one example of programs designed to promote preschool students' social-emotional development (Morris et al, 2014). K-12 programs, meanwhile, have started to make the development of non-cognitive skills an explicit focus. Many inner city charter schools, such as KIPP, Achievement First, and SEED Network schools, are investing in school cultures that emphasize strong work ethic, high self-efficacy, and dogged persistence in the face of failure in addition to the three R's (Seider, 2012; Tough, 2012; Whitman, 2008).

Institutions of higher education also recognize the importance of non-cognitive traits. DePaul University in Chicago, IL no longer requires potential applicants to submit the results of their performance on standardized tests like the SAT or ACT. Admissions officers instead evaluate potential students on the basis of short essay questions designed to capture students' non-cognitive skills. The University of Notre Dame Business School evaluates applicants with a student rating tool called the Personal Potential Index (PPI), developed by The Educational Training Service. The PPI emphasizes non-cognitive traits as well as academic ability, asking respondents to evaluate students on resilience and teamwork as well as knowledge, integrity, communication, and organizational skills.

While practitioners move forward with the implementation of programs and assessments focused on non-cognitive skills (Heitin, 2013), the research base lags behind. Non-cognitive

skills have proven elusive to measure (West, Kraft, Finn, Martin, Duckworth, Gabrieli & Gabrieli, 2014) and researchers must weigh the pros and cons of different modes of data collection. Surveys that ask students to self-rate are quick, cheap, and easy to administer but there is the potential for students to fake or exaggerate answers (Fan, Miller, Park, Winward, Christensen, Grotevant et al., 2006). This issue could become more salient as the stakes rise and if assessments of students' non-cognitive ability are incorporated into state accountability systems. A further problem arises when researchers attempt to compare the results of self-report surveys across different cultures and social contexts (Peng, Nisbett, & Wong, 1997). Fortunately, some promising solutions to this "reference bias" have been proposed (Hopkins & King, 2010; King, Murray, Salomon, & Tandon, 2004; King & Wand, 2007) and tested across a variety of fields (Kapteyn, Smith & Van Soest, 2007; Kristensen & Johansson, 2008; Pandey & Marlowe, 2014). Students can also be assessed by teacher ratings. Like self-reports, informant-reports are cheap and easy to administer but there is some evidence that teacher ratings may be confounded by students' demographic characteristics (Downey & Pribesh, 2004; Ehrenberg, Goldhaber, & Brewer, 1995; Ouazad, 2014). Although more time-consuming and expensive to administer, behavioral measures offer one potential avenue for unbiased assessments and may be more reliable than teacher ratings or self-report measures when individuals are unlikely, unwilling, or unable to report some things accurately (Fiske, 1978; Nisbett & Wilson, 1977). Moving forward, a comprehensive approach to measuring non-cognitive skills that incorporates multiple measures might be the most promising way to generate unbiased, efficient measures.

Students' non-cognitive skills may well deserve a place at the heart of education as a complement to traditional measures of academic achievement on standardized tests but they cannot assume this place without valid and reliable measures. In this study, we draw on data

from a site visit to a high-performing residential public school in Arkansas to test four potential measures of students' non-cognitive skills. In particular, we examine the relationship between student grit (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009), measured by a self-report survey instrument, and three behavioral measures of persistence, conscientiousness, and delay of gratification. Our specific research questions were as follows:

*Research Question 1:* Is self-reported student “grit” related to behavioral measures of persistence, conscientiousness, and delay of gratification measured concurrently?

*Research Question 2:* Are the three behavioral measures of students' non-cognitive skills related to each other?

*Research Question 3:* Do students' scores on any of the four non-cognitive measures vary by race, gender, grade, or religious background?

## **Method**

### *Sample*

Student participants in this study ( $N = 174$ , 48% male) were from a selective-admissions public boarding school in Arkansas that served approximately 240 11<sup>th</sup> and 12<sup>th</sup> grade students in school year 2012-13. The residential school is located in a small city in Arkansas with a population of approximately 35,000. The majority of our sample was in 11<sup>th</sup> grade (58%) at the time of data collection. Almost three-quarters of participants (71%) were White, 11% were Asian, 4% were African American, and the remainder was of another race/ ethnicity. A third of our sample (33%) was Protestant, 26% reported having no religious background, 10% were Catholic, and the remaining students belonged to other religions.

### *Procedure and Measures*

Nine researchers from the University of Arkansas collected these data on the school's campus over the course of a single school day in February 2013. Participants completed a preliminary activity designed to calculate a behavioral measure of individual persistence, which was followed immediately by a paper-and-pencil survey that consisted of 55 items.

The persistence activity employed was an unsolvable anagram measure, which has been widely used in numerous studies (eg. MacLeod et al., 2002; Muraven, Tice, & Baumeister, 1998; Smith, Kass, Rotunda, & Scheider, 2006; Toburen & Meier, 2010; White, Suway, Pine, Bar-Haim & Fox, 2011). The activity was presented to students as an untimed "challenging task." Students were given a set of ten anagrams and instructed to unscramble the letters so the words made sense. Only four of the ten anagrams were actually solvable (solvable: *tnkoe/token*, *yncfa/fancy*, *ruchs/crush*, *drnba/brand*; unsolvable *padus*, *alavt*, *dhboc*, *vaofea*, *kylix*, *malae*; taken from Toburen and Meier, 2010). Participants were told they could take as much time as they needed but they could stop any time they finished or decided to give up. The amount of time a student devoted to this task was recorded as the behavioral measure of persistence.

The student survey was administered immediately after the persistence activity. This study draws on three scales— a grit scale, an honesty scale, and a creativity scale— and general demographic information including gender, grade, race/ ethnicity, and religious background. Observed internal reliabilities were measured using Cronbach's coefficient alpha (Cronbach, 1951) for the grit, honesty, and creativity scales and were calculated as .72, .75, and .65, respectively. Only the grit scale was used to elicit a specific construct in the analysis, however.

As we describe below, student responses to the other two scales were used to calculate a behavioral measure of conscientiousness.

The 8-Item Grit Scale for Children is an adaptation of the Grit-S scale (Duckworth & Quinn, 2009) by the original authors, which has been shown to predict success measures such as educational attainment among adults, grade point average among Ivy League undergraduates, retention in the United States Military Academy, and ranking in the National Spelling Bee. Although highly correlated with Big Five Conscientiousness, Duckworth et al. (2007) found grit to have predictive validity of success measures beyond traditionally- defined conscientiousness or IQ. The grit scale for children uses a 5-point Likert-type scale ranging from 1 (*Not at all like me*) to 5 (*Very much like me*). Table 1 displays the descriptive statistics for this scale. The skewness and kurtosis values do not exceed |2|, implying there is no violation of the normality assumption.

<< TABLE ONE ABOUT HERE >>

The honesty scale used five total items— four items from Peterson and Seligman (2004) and the additional item, “I am honest,” chosen for its prima facie validity (Becker, 1998). The five items comprising the creativity scale were adopted from the KAI Inventory (Kirton, 1976). Both scales were measured on a Likert scale with the options *Strongly Agree*, *Agree*, *Disagree*, *Strongly Disagree*, and *Not Sure*. Responses to the honesty and creativity scales were used to generate a behavioral measure of conscientiousness by assessing the number of times a survey responder claimed ignorance to a self-reflective question that they should be able to answer without difficulty. The literature on item non-response has shown that failing to answer a survey question or claiming ignorance by selecting *Don't know* or *Not sure* is a reflection of both

cognitive and non-cognitive ability and predicts high school graduation (Hitt and Trivitt, 2013). Further, item non-response has been shown to predict earnings more strongly than traditional measures of either IQ or conscientiousness (Hedengren & Stratmann, 2012). Thus, we selected 11 self-reflective items from the honesty and creativity scales and recorded the frequency with which respondents claimed ignorance by selecting *Not Sure* (Table 2). This count-based variable serves as a second behavioral measure of students' non-cognitive ability.

<< TABLE TWO ABOUT HERE >>

The final behavioral measure we collected assessed an individual's ability to delay gratification by using the commonly used Delayed Payment Choice Task. This task, which is a widely used standard in the psychological literature, measures the ability of an individual to resist temptation now in exchange for a larger reward at a later time (Duckworth & Seligman, 2005; Metcalfe & Mischel, 1999; Vohs et al., 2008). Specifically, participants were given the choice of a \$5 reward immediately or a \$6 reward in one week for their participation in the study. They circled their preference on the last page of the survey packet and every student received a plain white envelope at the end of the experimental session that either contained \$5 or a slip of paper guaranteeing the student they would receive \$6 in one week's time. Monetary choice procedures like this have been previously shown to be valid with adolescents (Duckworth & Seligman, 2005; Reynolds, 2004) and to be a valid indicator of an individual's ability to delay gratification (Reynolds & Schiffbauer, 2005). A participant's choice of the delayed payment option formed our third behavioral measure of students' non-cognitive ability, coding the choice to take the payment immediately as "0" and the choice to wait as "1."

After survey administration, participants were probed for suspicion with a two question funneled procedure that asked them what they thought was the purpose of the study and if they thought any of the tasks were connected and if so, how. Finally, participants were debriefed through a letter that was distributed at the end of the school day, once all participants had completed the experiment.

## Results

The descriptive statistics for the three behavioral scales and the self-reported grit scale are presented in Table 3. On a scale from 1 to 5, the mean grit score was 3.16 ( $SD = .58$ ). Students spent an average of 705.54 seconds (almost 12 mins) on the unsolvable anagrams ( $SD = 361.97$ ). The maximum time allowed was 1,500 seconds (ie. 25 minutes), which 9 participants reached. The rate at which students claimed ignorance by selecting *Not Sure* was quite low— an average of 0.52 times. Finally, 58% of our sample demonstrated the ability to delay gratification by electing to wait a week for the higher payout of \$6 rather than take \$5 immediately.

<< TABLE THREE ABOUT HERE >>

*Research Question 1.* We calculate Pearson product-moment correlation coefficients to assess the strength of the relationship between students' score on the 8-Item Grit Scale for Children and the three behavioral measures of students' non-cognitive skills. We find no significant correlation between students' grit score and any of the three behavioral measures of students' non-cognitive ability (Table 4).

<< TABLE FOUR ABOUT HERE >>

*Research Question 2.* Table 4 also displays the Pearson product-moment correlation coefficients between the three behavioral measures. Although we find no relationship between the grit scale and any of the behavioral measures, we do find that two of the behavioral measures are related to one another. We observe a statistically significant relationship ( $r = .23, p < .05$ ) between the choice of delayed payment and time spent on the unsolvable anagrams (Table 4). We find no significant relationship between claiming ignorance to self-reflective questions (i.e. selecting *Not Sure*) and the amount of time a student was willing to spend on the unsolvable anagrams or between claiming ignorance and the choice of delayed payment, which is perhaps unsurprising given the small amount of variation in the Claiming Ignorance variable.

Given the significant relationship between students' choice of delayed payment and time spent on the unsolvable anagrams, we run a linear regression to better understand the relationship between these two variables. Specifically, we regress time spent on the unsolvable anagrams on the indicator for the choice of delayed payment, student race, student gender, and grade. Results are displayed in Table 5, with and without clustering standard errors by the researcher who administered the survey. A student who chose the delayed payment option spent 164 seconds (over 2 ½ mins) longer on the challenging task ( $p < .01$ ). These results suggest a strong relationship between delay of gratification and time spent on a challenging task.

<< TABLE FIVE ABOUT HERE >>

*Research Question 3.* We employ a multiple-sample generalization of the Mann-Whitney rank sum test—the Kruskal-Wallis one-way analysis of variance (Kruskal and Wallis, 1952, 1953)—to test for differences in grit scores by student race. We find a non-significant result,  $\chi^2(3, N= 174) = 2.75, p = .43$ , indicating that there is no relationship between student race and grit

scores. We employ the same procedure to evaluate differences in grit scores by students' religious background. Once again, we find a non-significant result,  $\chi^2(3, N=174) = 4.04, p = .26$ . We also test whether boys have higher grit scores than girls using an independent sample  $t$  test. Among 90 girls and 84 boys, girls ( $M = 3.21, SD = .05$ ) have higher average grit scores than boys ( $M = 3.10, SD = .07$ ), but the difference is not statistically significant ( $t(172) = 1.30, p = .19$ ). Finally, we evaluate differences in grit scores by student grade using an independent-sample  $t$  test. The result is statistically significant,  $t(172) = 1.85, p = .07$ . Comparing 73 12<sup>th</sup> grade students and 101 11<sup>th</sup> grade students, the 12<sup>th</sup> grade students had higher average grit scores ( $M = 3.25, SD = .07$ ) than their younger counterparts in the 11<sup>th</sup> grade ( $M = 3.09, SD = .05$ ).

We also test for differences in the three behavioral measures by race, religious background, gender, and grade. We find significant differences between 11<sup>th</sup> and 12<sup>th</sup> grade students on all three behavioral measures. Regarding delay of gratification, 12<sup>th</sup> grade students are less likely to wait a week for the higher payout ( $M = .49, SD = .06$ ) than their 11<sup>th</sup> grade counterparts ( $M = .64, SD = .05$ ),  $t(172) = 1.20, p = .05$ . Twelfth grade students spend less time on the unsolvable anagrams ( $M = 651.16, SD = 34.70$ ) than 11<sup>th</sup> grade students ( $M = 751.14, SD = 39.52$ ),  $t(172) = 1.82, p = .07$ . Twelfth grade students were also more likely to claim ignorance by selecting *Not Sure* ( $M = .71, SD = .14$ ) than their 11<sup>th</sup> grade counterparts ( $M = .36, SD = .07$ ),  $t(172) = 2.48, p = .01$ .

Finally, we observe a significant difference by race in the amount of time spent on the unsolvable anagrams,  $\chi^2(3, N=174) = 7.45, p = .06$ . We employ post-hoc tests to evaluate pairwise differences among the four racial categories (White, Asian, Black, Other), controlling for Type I error with a Bonferroni correction. The results of these tests indicate a significant difference between Asian and Black students ( $z = 2.41, p = .008$ ) and between Asian and White

students ( $z = 3.14, p = .001$ ), with Asian students spending longer on the unsolvable anagrams than both other groups.

## **Discussion**

This study reveals a number of key insights for researchers and practitioners. Contrary to expectations, we find that a student's self-reported grit score is uncorrelated with behavioral measures of their persistence, conscientiousness, and ability to delay gratification. There are two possible explanations for this. The first is that the four constructs we measure do not overlap sufficiently. Grit is defined as "passion and perseverance for long-term goals" (Duckworth et al., 2007) whereas our behavioral measures capture persistence on a challenging task, conscientiousness, and the ability to delay gratification for a short period of time. It could be the case that the four constructs we are attempting to measure are simply orthogonal to one another. This explanation would be plausible if none of the behavioral measures were related to one another either but is disproved by the results of Research Question 2, which finds a significant correlation between persistence (i.e. time spent on the unsolvable anagrams) and delay of gratification (i.e. selecting the later payout). A student who chose the delayed payment option spent 164 seconds (over 2 ½ mins) longer on the challenging task. A more plausible explanation therefore, is that there are measurement problems with all four measures, but particularly with the grit scale. Duckworth et al. (2007) acknowledge that the grit scale relies upon self-report, making it highly susceptible to social desirability bias; that is, systematic error in responses resulting from a desire by participants to project a favorable image (Maccoby & Maccoby, 1954). Prior research has shown that this type of response bias can increase measurement error (Cote & Buckley, 1988) and shrink, inflate, or moderate relationships between variables (Zerbe & Paulhus, 1987).

It is also interesting to note that the grit results appear to contradict the behavioral results. That is, 11<sup>th</sup> grade students rate themselves lower on self-reported grit but outperform their 12<sup>th</sup> grade counterparts on the three behavioral measures of persistence, delay of gratification, and conscientiousness. How do we make sense of this? A plausible explanation is that the 11<sup>th</sup> grade students, who exhibit high levels of persistence and delay of gratification on the behavioral measures, may be highly self-critical on the self-reported grit scale. This would explain why self-reported grit is underestimated. The 11<sup>th</sup> grade students in this sample may be especially susceptible to this type of bias because they are new to this high-performing school that draws students from all over the state and are still adjusting to being surrounded by high-achieving peers. This type of reference bias commonly arises in cross-cultural comparisons (Peng, Nisbett, & Wong, 1997) and researchers have suggested numerous solutions to address it (Hopkins & King, 2010; King et al, 2004; King & Wand, 2007). In addition to incorporating “anchoring vignettes” into student surveys, future research might address this problem by administering behavioral measures of non-cognitive ability, which are less subject to manipulation by respondents.

The policy implications of this research are timely. While individual schools such as those in the KIPP charter school network are already experimenting with subjective character report cards similar to traditional academic report cards, it would be wise for state accountability systems to postpone mandating the incorporation of such measures into high stakes decisions until we know more about the performance of different measures under a variety of circumstances. This study contributes in one way by documenting how four measures perform in a high-performing, selective admissions public school environment.

## **Conclusion**

With an abundance of increasingly available datasets of students' standardized test score data, researchers and education policymakers run the risk of rating student development by focusing exclusively on cognitive outcomes in key academic subjects such as math and reading. Momentum is building among practitioners, however, to develop and measure students' non-cognitive skills. Despite popular interest in this topic, our knowledge of how to measure these skills is still under-developed. This study provides important information on the strengths and weaknesses of four measurement tools readily available to researchers and practitioners.

More research is needed to develop and test new measures of non-cognitive ability so that practitioners can access an entire toolbox of measures. Better instruments will allow us to identify and support effective interventions and reforms designed to foster the development of non-cognitive skills, allowing educators to develop programs aimed at students at risk of dropping out or failing to meet other important educational milestones such as high school graduation and college completion.

## References

- Becker, T.E. (1998). Integrity in organizations: Beyond honesty and conscientiousness, *The Academy of Management Review*, 23(1): 15-161.
- 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.
- Cote, J. A. & Buckley, M. R. (1988). Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation, *Journal of Consumer Research*, 14 (4), 579- 582.
- Cunha, F., & Heckman, J.J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738–782.
- Daily, S., Burkhauser, M. & Halle, T. (2010). A review of school readiness practices in the states: Early learning guidelines and assessments, *Child Trends*, 1(3). Retrieved from <http://www.childtrends.org/wp-content/uploads/2013/05/2010-14-SchoolReadinessStates.pdf>
- DeRidder, D.T.D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F.M., & Baumeister, R.F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors, *Personality and Social Psychology Review*, 16(1): 76-99.
- Dee, T.S., & West, M.R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, 33(1): 23–46.
- Downey, D.B., & Pribesh, S. (2004). When race matters: Teachers’ evaluations of students’ classroom behavior. *Sociology of Education*, 77(4): 267–282.
- 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. & Seligman, M.E.P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents, *Psychological Science*, 16(12): 939- 944.
- 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), 1087–1101.
- Duckworth, A.L., Tsukayama, E. & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control, *Social Psychological & Personality Science*, 1(4): 311-317.

- Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D., & Schellinger, K.B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432.
- Ehrenberg, R.G., Goldhaber, D.D., & Brewer, D.J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Education Longitudinal Study of 1988. *Industrial and Labor Relations Review*, 48(3), 547–561.
- Fan, X., Miller, B.C., Park, K., Winward, B.W., Christensen, M., Grotevant, H.D. & Tai, R.H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods*, 18(3): 223–244.
- Fiske, D.W. (1978). *Strategies for personality research*, San Francisco: Jossey-Bass.
- Goldstein, J., Eastwood, M., & Behuniak, P. (2014). Can teacher ratings of students' skills at kindergarten entry predict kindergarten retention, *The Journal of Educational Research*, 107(3): 217-229.
- Hendengren, D. & Stratmann, T. (2012). The dog that didn't bark: What item non-response shows about cognitive and non-cognitive ability. Working paper. Retrieved from [http://econlog.econlib.org/archives/2012/11/hedengrens\\_dog.html](http://econlog.econlib.org/archives/2012/11/hedengrens_dog.html)
- Heitin, L. (2013, October 16). Teachers use social-emotional programs to manage classes. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2013/10/16/08social.h33.html?qs=sel>
- Hitt, C. & Trivitt, J. (2013). Don't know? Or don't care? Predicting educational attainment using survey item response rates and coding speed tests as measures of conscientiousness. *EDRE Working Paper No. 2013-05*. Retrieved from <http://www.uaedreform.org/dont-know-or-dont-care-predicting-educational-attainment-using-survey-item-response-rates-and-coding-speed-tests-as-measures-of-conscientiousness/>.
- Hopkins, D.J., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2), 201–222.
- Jackson, C.K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. *NBER Working Paper No. 18624*. Retrieved from <http://www.nber.org/papers/w18624>.
- Kapteyn, A., Smith, J.P., & Van Soest, A. (2007). Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, 97(1): 461–73.
- King, G., & Wand, J. (2007). Comparing incomparable survey responses: New tools for anchoring vignettes. *Political Analysis*, 15(1), 46–66.
- King, G., Murray, C., Salomon, J.A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191–207.

- Kirton M.J. (1976) Adaptors and Innovators: A description and a measure. *Journal of Applied Psychology*, 61(5): 622-629.
- Kristensen, N., & Johansson, E. (2008). New evidence on cross- country differences in job satisfaction using anchoring vignettes. *Labour Economics*, 15(1): 96–117.
- Kruskal, W.H., & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260): 583–621.
- Kruskal, W.H., & Wallis, W.A. (1953). Errata: Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 48: 907–911.
- Lucas, R. E., & Baird, B.M. (2006). Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29–42). Washington, DC: American Psychological Association.
- Maccoby, E.E. & Maccoby, N. (1954). The Interview: A Tool of Social Science. In L. Gardiner (Ed.), *Handbook of Social Psychology* (Vol. 1, pp. 449-487). Cambridge, MA: Addison-Wesley.
- MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal bias of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, 111(1): 107-123.
- Metcalf, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower, *Psychological Review*, 106(1): 3-19.
- Mischel, W., Shoda, Y., & Peake, P.K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, 54(4): 687-696.
- Mischel, W., Shoda, Y., & Rodriguez, M.L. (1989). Delay of gratification in children. *Science*, 244(4907), 933–938.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2693–2698.
- Morris, P., Mattera, S.K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence* (OPRE Report No. 2014-44). Washington, D.C.: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from [http://www.mdrc.org/sites/default/files/HSCares%20Impact\\_ExecSummary%20MDRC.pdf](http://www.mdrc.org/sites/default/files/HSCares%20Impact_ExecSummary%20MDRC.pdf).

- Muraven, M., Tice, D.M., & Baumeister, R.F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74(3): 774-789.
- Nisbett, R. E. & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3): 231-259.
- Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance and Policy*, 9(3), 334–372.
- Pandey, S.K., & Marlowe, J. (2014). Assessing survey-based measurement of personnel red tape with anchoring vignettes. *Review of Public Personnel Administration*. Retrieved from <http://rop.sagepub.com/content/early/2014/04/20/0734371X14531988.abstract>.
- Peng, K., Nisbett, R.E., & Wong, N.Y.C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods*, 2(4), 329–344.
- Peterson, C., & Seligman, M.E.P. (2004). *Character strengths and virtues: A handbook and classification*. Washington DC: American Psychological Association.
- Reynolds, B., & Schiffbauer, R. (2005). Delay of gratification and delay discounting: A unifying feedback model of delay-related impulsive behavior. *Psychological Record*, 55(3): 439–460.
- Reynolds, B. (2004). Do high rates of cigarette consumption increase delay discounting? A cross-sectional comparison of adolescent smokers and young-adult smokers and nonsmokers. *Behavioural Processes*, 67(3): 545–549.
- Seider, S. (2012). *Character Compass: How Powerful School Culture Can Point Students Toward Success*. Cambridge, MA: Harvard Education Press.
- Shoda, Y., Mischel, W., & Peake, P. (1990) Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, 26(6): 978-986.
- Smith, S.A., Kass, S.J., Rotunda, R.J., & Schneider, S.K. (2006). If at first you don't succeed: Effects of failure on general and task-specific self-efficacy and performance. *North American Journal of Psychology*, 8(1): 171-182.
- Stouffer, S. (1955). *Communism, conformity, and civil liberties*. New York: Doubleday.
- Toburen, T. & Meier, B.P. (2010). Priming God-related concepts increases anxiety and task persistence, *Journal of Social and Clinical Psychology*, 29(2), 127-143.
- Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. Houghton Mifflin Harcourt.

- Vohs, K.D., Baumeister, R.F., Schmeichel, B.J., Twenge, J.M., Nelson, N.M., & Tice, D.M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, 94(5): 883-898.
- West, M. R., Kraft, M. A., Finn, A. S., Martin, R., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Working Paper*. Retrieved from <http://www.gse.harvard.edu/cepr-resources/files/news-events/cepr-promise-paradox.pdf>
- White, L.K. Suway, J.G., Pine, D. S., Bar-Haim, Y. & Fox, N.A. (2011) Cascading effects: The influence of attention bias to threat on the interpretation of ambiguous information. *Behaviour Research and Therapy*, 49(4): 244-251.
- Whitman, D. (2008). *Sweating the small stuff: Inner-city schools and the new paternalism*. Washington D.C.: Thomas B. Fordham Institute Press.
- Wulfert, E., Block, J.A., Santa Ana, E., Rodriguez, M.L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, 70(4), 533–552.
- Zerbe, W.J. & Paulhus, D.L. (1987), Socially Desirable Responding in Organizational Behavior: A Reconception, *Academy of Management Review*, 12 (2), 250-264.
- Zimmerman, B.J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91.

Table 1. Item Content and Descriptive Statistics for the Grit Scale for Children

	M	SD	Skewness	Kurtosis
New ideas and projects sometimes distract me from previous ones	3.53	0.82	-.24	0.16
Setbacks (delays and obstacles) don't discourage me	3.13	1.01	.05	-0.72
I have been obsessed with a certain idea or project for a short time but later lost interest	3.42	1.07	-.18	-0.76
I am a hard worker	3.82	.90	-.58	0.08
I often set a goal but later choose to pursue (follow) a different one	2.98	.97	.01	-0.62
I have difficulty maintaining (keeping) my focus on projects that take more than a few months to complete	3.13	1.16	-.04	-0.91
I finish whatever I begin	3.58	.92	-.15	-0.60
I am diligent (hard working and careful)	3.80	.93	-.41	-0.47

Note.  $N= 174$

Table 2. Item Content Used to Generate Behavioral Measure of Conscientiousness

Item	Scale of Origin	Percentage of Respondents Selecting "Not Sure"
I always keep my word	Honesty	0.00
I lie to get myself out of trouble	Honesty	2.30
I tell the truth, even if it gets me in trouble	Honesty	4.60
I lie to get what I want	Honesty	5.17
I am honest	Honesty	8.05
I have a lot of creative ideas	Creativity	2.30
I prefer tasks that enable me to think creatively	Creativity	4.02
I often take risks doing things differently	Creativity	2.87
I like to do things in an original way	Creativity	5.75
I have fresh perspectives on old problems	Creativity	15.52

Note:  $N = 174$

Table 3. Descriptive Statistics of All Conscientiousness Measures

Measure	Construct	Format	<i>M</i>	<i>SD</i>	Min	Max
1. 8-Item Grit Scale for Children	Grit	Self-Report	3.16	.57	1.75	4.63
2. Time on Unsolvable Anagrams	Persistence	Behavioral	709.20	360.82	105	1500
3. Claiming Ignorance	Conscientiousness	Behavioral	.51	.95	0	6
4. Delayed Payment	Delay of Gratification	Behavioral	.58	.49	0	1

Note: *N* = 174. Time on unsolvable anagrams refers to the number of seconds voluntarily spent attempting to unscramble letters to make words; Claiming ignorance refers to selecting *Not Sure* to a self-reflection survey question; Delayed payment refers to waiting one week for \$6 in compensation instead of accepting \$5 immediately.

Table 4.

*Correlations among four behavioral measures of persistence*

	1.	2.	3.	4.
1. Time on Unsolvable Anagrams	1.00			
2. Claiming Ignorance	-.07	1.00		
3. Choice of Delayed Payment	.23**	-.05	1.00	
4. Grit Scale	.02	-.02	-.01	1.00

Note.  $N= 174$ ; \*\*  $p < 0.05$

Table 5.

*Relationship between Persistence and Delay of Gratification*

	(1)	(2)
Delayed Payment	165.84** (55.34)	165.84** (55.56)
Male	-2.25 (55.93)	-2.25 (61.09)
Asian	114.59 (88.60)	114.59* (56.44)
Black	115.39 (142.07)	115.39 (128.30)
Other Race	15.30 (83.60)	15.30 (72.76)
Catholic	-36.03 (100.88)	-36.03 (140.48)
Other Religion	-107.37 (69.81)	-107.37* (56.04)
No Religion	-109.41 (72.17)	-109.41* (57.02)
12 <sup>th</sup> Grade	-96.81* (56.12)	-96.81* (49.99)
Observations	174	174
R-Squared	.09	.09
Clustered Standard Errors	No	Yes

Note: \*  $p < 0.10$ , \*\*  $p < 0.05$ . The dependent variable is a measure of persistence, calculated as the amount of time, in seconds, spent working on unsolvable anagrams. Standard errors in parentheses. Omitted religious group is Protestant; Omitted race is White.