

University of Arkansas, Fayetteville

ScholarWorks@UARK

---

Education Reform Faculty and Graduate  
Students Publications

Education Reform

---

10-2018

## Further Validation of Survey Effort Measures of Relevant Character Skills: Results from a Sample of High School Students

Gema Zamarro

*University of Arkansas, Fayetteville*

Malachi Nichols

*University of Arkansas, Fayetteville*

Angela Duckworth

*University of Pennsylvania*

Sidney D'Mello

*University of Colorado Boulder*

Follow this and additional works at: <https://scholarworks.uark.edu/edrepub>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Education Policy Commons](#), and the [Secondary Education Commons](#)

---

### Citation

Zamarro, G., Nichols, M., Duckworth, A., & D'Mello, S. (2018). Further Validation of Survey Effort Measures of Relevant Character Skills: Results from a Sample of High School Students. *Education Reform Faculty and Graduate Students Publications*. Retrieved from <https://scholarworks.uark.edu/edrepub/66>

This Article is brought to you for free and open access by the Education Reform at ScholarWorks@UARK. It has been accepted for inclusion in Education Reform Faculty and Graduate Students Publications by an authorized administrator of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu).



UNIVERSITY OF  
ARKANSAS

College of Education & Health Professions  
*Education Reform*

## **WORKING PAPER SERIES**

### **Further Validation of Survey Effort Measures of Relevant Character Skills: Results from a Sample of High School Students**

Gema Zamarro, Malachi Nichols, Angela L. Duckworth, Sidney K. D'Mello

October 2018

EDRE Working Paper 2018-07

The University of Arkansas, Department of Education Reform (EDRE) working paper series is intended to widely disseminate and make easily accessible the results of EDRE faculty and students' latest findings. The Working Papers in this series have not undergone peer review or been edited by the University of Arkansas. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as an EDRE working paper.

## Further Validation of Survey Effort Measures of Relevant Character Skills: Results from a Sample of High School Students

**Gema Zamarro\***

University of Arkansas

**Malachi Nichols**

University of Arkansas

**Angela L. Duckworth**

University of Pennsylvania

**Sidney K. D'Mello**

University of Colorado Boulder

First Version: February 2017

This Version: October 2018

### **Acknowledgments**

We would like to thank Julie Trivitt for her help in the early stages of this paper and Albert Cheng and Collin Hitt for their comments and feedback on our results. We also thank conference participants at the 42<sup>nd</sup> AEFPP Annual Conference and the University of Arkansas Department of Education Reform Brownbag Seminar Series for all their feedback. Any remaining errors are our own.

\* Corresponding author. The 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](mailto:gzamarro@uark.edu).

## **Abstract**

Character skills, including conscientiousness, grit or self-control are important determinants of relevant life outcomes. However, researchers struggle to find valid measures of these skills and many existing datasets lack any measures of them at all. This limits research on how these important skills could be better supported and developed. Recent research has shown the potential of parametrizations of survey effort measures as proxy measures of character skills related to conscientiousness, to either complement other collected measures or to add to datasets that lack such measures. This study provides further validation of these survey effort measures in a sample of high school students by studying their relationship with external measures reported by teachers, other direct performance task measures of related skills, high school academic outcomes, and college attendance. Our results show promise for survey effort measures to be used as proxy measures of character skills related to grit and self-control.

**Keywords:** Survey Effort, Character Skills, Teacher Reports, High School Performance

**JEL codes:** C83, C91

## 1. Introduction

Though the importance of character skills, such as conscientiousness, grit or self-control, to life outcomes like education, labor, health outcomes, or criminal behavior is well-established (see, Almlund et al., 2011; Weel, 2008; Heckman, Stixrud, and Urzua, 2006), researchers have struggled to find valid measures of these skills with many existing datasets lacking any measures at all. As a result, research on how these important skills could be better supported and developed is limited.

Recent literature has proposed the use of survey effort measures as proxy measures of character skills related to conscientiousness to either supplement information obtained through self-reports, which might be affected by multiple types of bias (e.g. reference group bias and social desirability bias; see Duckworth and Yeager, 2015), or to complement datasets that lack any measures of these skills (Marcus and Schütz, 2005; Hitt, 2015; Huang et al., 2012; Johnson, 2005; Meade and Craig, 2012; Zamarro et al., 2018). The idea behind this approach is that surveys take the effort to complete. For students, in particular, surveys administered in schools could resemble schoolwork or homework. Therefore, by studying how much effort students put forward in surveys, we could obtain proxy measures of relevant character skills related to conscientiousness. Two parametrizations of survey effort measures that have shown promise as relevant proxies include: item non-response rates and careless answering.

Item non-response is defined as the percentage of questions skipped by a respondent on a survey. Taking advantage of longitudinal nationally representative samples of adolescents and adults from the United States and Germany, Hedengren and Stratmann (2012) find item non-response to be correlated with self-reported conscientiousness (one standard deviation increase in response rates was associated with a significant 0.3 standard deviations increase in self-reported conscientiousness) and a significant predictor of earnings and mortality risks. Furthermore, Hitt, Trivitt and Cheng (2016) use six different longitudinal nationally representative samples of American youth to determine the relationship between the percentage of questions skipped and desirable self-reported outcomes measured in adulthood. They found that item non-response was a significant predictor of self-reported educational attainment and labor market outcomes, independent of available cognitive ability measures (one standard deviation increase in item non-

response was associated with reporting completing between 0.1 and 0.3 fewer years of education).

In addition, some respondents might show low survey effort by answering randomly and overall carelessly (Johnson, 2005; Meade and Craig, 2012). This form of survey effort can be captured through parametrizations of careless answering. Using two national longitudinal surveys, Hitt (2015) found that careless answering in adolescent respondents was associated with reporting fewer years of final education and a decreased probability of high school completion, independent of cognitive ability (one standard deviation increase in careless answering was associated with an about 0.1 decrease in self-reported completed years of education and almost a two percentage points decrease in the probability of graduating from high school). Similarly, using data of a nationally representative internet panel of American adults, Zamarro et al. (2018) found that repeated careless answering behavior was negatively correlated with self-reported grit and self-reported conscientiousness (partial correlations ( $r_{xy,z}$ ) of about -0.15 after controlling for cognitive ability and demographic information), and positively correlated with neuroticism, shedding light on its construct validity. They also determined that careless answering is a significant negative predictor of self-reported final years of education and lower levels of self-reported income and labor outcomes.

Although recent research has shown that survey effort measures can be promising proxy measures for conscientiousness and related skills (Hedengren and Stratmann, 2012; Hitt, Trivitt and Cheng 2016; Hitt, 2015; Zamarro et al., 2018), these validation exercises have relied on self-reported measures of character skills and outcome variables to compare with, and lacked external sources of information<sup>1</sup>. We aim to fill this gap by studying the relationship between survey effort measures and teacher evaluations of skills, performance task measures, and external outcome measures.

We use data on high school seniors attending a public school in the Northeastern United States. Our dataset contains a rich set of information on different measures of students' character skills including self-reported measures, teacher's reports, and performance task measures from two validated tasks. We use this data to complement the work of Hitt, Trivitt and Cheng (2016)

---

<sup>1</sup> An exception is the work of Hedengren and Stratmann (2012) who used information on earnings and mortality risks from administrative sources.

and Hitt (2015) by studying the correlation of survey effort measures with students' self-reported measures of different character skills and academic outcomes at the end of high school, college attendance one year after graduation but, more importantly, with reports on these character skills provided by teachers. Secondly, we study the relationship between survey effort measures and other performance task measures designed to capture academic diligence (Galla et al., 2014) and frustration tolerance (Meindl et al., in press) that our high school seniors were asked to perform. Our results suggest that survey effort measures can be used as proxy measures of character skills related to grit, that is the tendency to sustain passion and interest for long-term goals, and self-control, that is the ability to control one's emotions, impulses, or the expression of them in one's behavior.

## **2. DATA**

The data used is part of a study on college persistence led by a research team at the University of Pennsylvania. In the spring of 2014, the research team collected data from 513 high school seniors attending a public high school in the Northeast of the United States. One year later, the team used the National Student Clearinghouse to track these students' college enrollment status. For more details on recruitment and data collection see Meindl et al. (in press).

### **2.1 Character Skills Measures**

#### ***2.1.1. Survey Effort Measures***

##### *Item Non-response*

The first way we parametrize survey effort is through measures of survey item non-response, following Hitt, Trivitt and Cheng (2016). We computed two measures with this information. Item non-response, which captures continuous item non-response rates, was computed by dividing the total number of questions a student leaves blank by the number of answerable questions to which a student should have responded, given legitimate skips. Relatedly, we also computed a dichotomous item non-response measure which is just a binary indicator taking the value of one if the student leaves any answerable question blank, dependent on legitimate skips. This was done because the sample was almost evenly divided with respect to the proportion of students who completed the entire survey, 47 percent. Figure 1 shows the distribution of survey item non-response rates in our sample.

### *Careless Answering*

The second way we parametrize survey effort is through measures of careless answering, following Hitt, (2015). The idea behind this measure is as follows. Consider a reliable, validated scale with a number of items. If the scale is reliable, each item will consistently measure the same underlying construct. Individual responses to each item would be well predicted by the responses to other items in this same reliable scale. Thus, we interpret deviations in responses from predicted values, given responses to other items in the scale, as measures of careless answering.

In practice, we first identify reliable scales that students had to answer, with Cronbach alpha reliability coefficients (Cronbach and Meehl, 1955) of at least 0.7. We exclude self-reported scales of grit and self-control used to validate survey effort measures in this paper. In total, we identified the following seven scales satisfying these conditions: a trust scale, a belonging scale, an interest in school scale, an academic self-efficacy scale, a distress tolerance scale, a purpose scale, and a brief self-control scale. For each scale, we regress responses from each item in the scale on the average score of the rest of the items in that scale. Third, residuals from each of these regression models are obtained to capture the extent to which the response to a particular item is unpredictable, based upon the response patterns observed by the current student and others in the analytic sample. 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 averaged residuals at the student level with higher values of this measure indicating higher levels of carelessness or unpredictability in responding.

Figure 2 shows the distribution of careless answering in our sample. Since the careless answering measure is standardized by construction, mean and standard deviation are not very informative. We observe, however, considerable variation across students in the sample.

Item non-response and careless answering appear to be different approaches to exert low survey effort. Within a given item in the survey, careless answering and item non-response are mutually exclusive approaches and so, it is impossible to observe both behaviors at the same



time. Over the course of the survey, students also appear to choose one of the approaches more frequently and few seem to switch back and forth between these behaviors, as the estimated correlation coefficient between these two measures of survey effort is only 0.17.

### ***2.1.2. Teacher Reports***

Three teachers (homeroom, English, and social studies) rated each student on a grit and self-control scale and answered additional questions about classroom behavior and work ethic. To assess teacher's reported grit levels of their students, teachers were shown the items, from the grit scale adapted from Duckworth et al. (2007), students were asked to complete, and were asked to rate how much these items as a whole described each student using a five-point Likert scale. Similarly, teachers were also asked to report on the level of self-control of their students, by overall evaluating the level of each student according to the eight-item scale adapted from the Domain-Specific Impulsivity Scale for Children (DSIS; Tsukayoma et al., 2013) students were asked to complete. Since three teachers reported on a single child, the individual scores were averaged for each student to give that student a unique construct score to increase validity (Eid and Diener, 2006). High scores represent higher levels of that student character skill.

To measure classroom behavior and work ethic, teachers were also asked to report on students' redirection and homework completion. For redirection, teachers were asked to pick the number of times the student required redirection within the last week. Reporting options included: zero, one, two, three, four, or five or more times. Teachers were instructed to remember that redirection includes reminding the student to stay on task or to follow classroom rules. Teachers' reported redirection times were averaged for each student to give each student a redirection score. A high number of redirects could represent a lack of diligence or negative classroom behavior. Finally, teachers were also asked about homework completion. In particular, teachers were asked what percentage of assignments, from zero to one hundred, did the student complete on time and received a passing grade. The teachers' reported values were then averaged to give the respective student an average percentage of times he/she successfully completes homework on time. A higher percentage infers that the student has high levels of work ethic.

### ***2.1.3. Direct Performance Task Measures***

### *Academic Diligence Task (ADT)*

The ADT is a computer-generated task designed to measure academic diligence (Galla et al., 2014). In this computer task, students are given the option to perform simple math problems, after being prompted of the benefits of this type of exercises, or consume media, by watching online clips or playing online games. We measure academic diligence by the percentage of the total task time, 12 minutes, a student spent completing math problems instead of consuming media or playing games. Higher percentages represent higher levels of student's academic diligence.

Using a sample of over 900 high school students, Galla et al. (2014) found that measures of student engagement through the ADT task were correlated with student self-reported measures of conscientiousness ( $r_{xy,z} = 0.09$ ), self-control ( $r_{xy,z} = 0.15$ ), and grit ( $r_{xy,z} = 0.17$ ). Also, performance in the ADT was predictive of student's high school Grade Point Average (GPA), standardized test scores, high school graduation, and college enrollment, even after controlling for potential confounds including cognitive ability<sup>2</sup>.

### *Mirror Tracing Frustration Task (MTFT)*

Students in our sample were also asked to complete the Mirror Tracing Frustration Task (MTFT; Meindl et al., in press). The MTFT measures frustration tolerance, which is a person's ability to persist through tasks that cause frustration and appear out of their control. During this task, students were given the option to trace a shape using the mouse on their computer or consume media by watching online videos. However, the task was such that actual actions with the mouse produced movements in the opposite direction. There was also a random drift added to each mouse moment, so perfect control was not possible. This required students to present high levels of concentration when performing the task and induced frustration. If the student stopped tracing or traced off of the shape, the task automatically restarted. Students had the option to switch between the task and media as often as they desired but were informed about the importance of developing perceptual-motor skills for various real-world tasks. Frustration tolerance was measured as the percentage of the total task assigned time, 5 minutes, a student

---

<sup>2</sup> The convergent and incremental predictive validity of the ADT task was replicated and confirmed within our sample. Results are available from the authors upon request.

spent tracing. Using this same data Meindl et al. (in press) showed that higher frustration tolerance was significantly associated with self-reported and teacher reported grit and self-control measures ( $r_s = 0.11$  to  $0.22$ ), as well as high school GPA, standardized test scores, and college persistence.

#### ***2.1.4. Self-Reported Measures***

For completeness, we also study the relationship between survey effort measures and the following self-reported measures collected in the study.

##### *Grit*

Students were asked to complete a five-item scale adapted from the full grit scale (Duckworth et al., 2007). Students were asked to rate how true five statements described themselves on a five-point Likert-type scale (*Not at all true; Slightly true; Somewhat true; Very true; Completely true*). These statements included, for example, “I finish whatever I begin” and “I stay committed to my goals.” Each student’s question scores were averaged to develop a grit score for each respondent. Possible grit scores range from one to five with a high score representing high values of grit. This scale showed high reliability in our sample with a Cronbach alpha measure of 0.8.

##### *Self-Control*

Students were also asked to complete a self-control questionnaire adapted from the DSIS scale (Tsukayoma et al., 2013) described above. This scale consisted of a combination of four questions pertaining to schoolwork and four questions pertaining to interpersonal situations. The scale asked students to rate how true the eight statements described themselves on a five-point Likert-type scale (*Not at all true; Slightly true; Somewhat true; Very true; Completely true*). The statements for work skills included questions like “I come to class prepared” and “I get to work right away, instead of waiting until the last minute”, while the statements relating to interpersonal skills included questions like “I allow others to speak without interruption” and “I control my temper.” Scores from each scale were averaged to develop a self-control combined score for each student. Average scores were also computed separately to represent self-control in work and self-control related to interpersonal skills. Scores range from 1 to 5 with a high score

meaning the student has high levels of self-control. The self-control combined scale showed high reliability in our sample with a Cronbach alpha measure of 0.8.

## **2.2. Outcome Measures**

We also study the relationship between survey effort measures and other available outcome variables, as a way to complement prior work validating these measures. For this purpose, our outcome measures include high school GPA ranging from zero to 100, a binary variable indicating if the student graduated high school, a binary variable indicating if the student attempted to take the SAT, the total SAT score in the first attempt, for those who attempted the test, ranging from 600 to 2400, including the sum of the SAT critical reading, SAT math, and SAT writing scores. Furthermore, we constructed three binary variables indicating if the student was continuously enrolled in any college for one year after graduating high school, indicating if the student was continuously enrolled in a four-year college one year following high school graduation, and indicating if the student was continuously enrolled full-time in a four-year college one year after high school. Additionally, we also study the relationship between survey effort measures and performance on students' end of the year assessments in math and reading. Results in these exams are part of students' state graduation requirements and scores range from 1200 to 1800.

## **2.3. Cognitive Ability and Other Relevant Information Available for our Analysis**

We would like to find that our measures of survey effort are predictive of character skills and outcomes after controlling for cognitive ability. For this purpose, we have measures of students' performance on the matrix reasoning subset of the Kaufman Brief Intelligence Test (KBIT; Kaufman and Kaufman, 1990), with scaled scores ranging from 40 to 132.

Finally, our analysis also includes controls for students' socio-demographic information such as age, gender, ethnicity, English Language Learner (ELL), Special Education Status (SPED), Free and Reduced Lunch (FRL) status, and a measure of household income.

## **3. Empirical Strategy for Validation of Measures**

For survey effort measures to be valid proxy measures of conscientiousness, they should be correlated with other measures of similar character skills (convergent validity) as well as with other outcome variables known to be correlated with the same latent skills (criterion validity).

Accordingly, we computed Spearman correlations and partial rank correlations<sup>3</sup> (controlling for cognitive ability and socio-demographic information) for our measures of survey effort, i.e. non-response rates and measures of careless answering, and self-reported measures of grit and self-control. We expect to find negative correlations of both survey effort measures with these self-reports. We also expect to find negative correlations between measures of survey effort and teachers' reported self-control, grit and homework completion and positive correlations with redirection. Finally, we expect to find negative correlations between survey effort, diligence and frustration tolerance, as measured through these performance tasks.

The last set of analyses looks at criterion validity of survey effort measures. To do so, we estimate linear regression models and linear probability models<sup>4</sup> to predict each of the following academic outcomes: high school GPA, high school graduation, attempted to take the SAT, SAT scores if attempted, end of the year math and reading test scores, college enrollment in the first year, college enrollment in a four-year college and full-time college enrollment in a four-year college. We estimate separate models for item non-response rates, dichotomous non-response, and careless answering measures, following this type of specification:

$$Academic\ Outcome_i = \beta_0 + \beta_1 Survey\ Effort_i + \beta_2 Cognitive\ Ability_i + \beta_3 X_i + \varepsilon_i \quad (1)$$

Our models control for cognitive ability measures through the use of the KBIT scaled score.  $X_i$  represents a vector of student socio-demographic controls including age, ethnicity, gender, ELL status, FRL status, SPED status, and parental income. We report standardized regression coefficients for all models. For comparison, we also estimate models including direct performance task measures of academic diligence and frustration tolerance.

---

<sup>3</sup> Variables were first ranked and partial correlations were then obtained using the command *pcorr* in STATA.

<sup>4</sup> For binary outcomes, we also estimated discrete choice logit models. Results were similar to those of linear probability models presented here. We present the results of standardized coefficients from linear probability models here to ease the comparison of effects for different outcome variables. Results from logit models are available from the authors upon request.

## 4. Results

### 4.1 Descriptive Statistics

Table 1 presents descriptive statistics for our sample of 513 high school seniors. The sample is about equally split among male and female students. Most of the students in our sample are African American, White, or Asian in proportions of 41, 36 and 20 percent, respectively. Only three percent of the students in our sample were Hispanic. Students in our sample reside in homes with an average household income of about \$52,000. Also slightly more than half of our sample qualifies for free and reduced lunch. These results make this sample relatively more African American and lower income than the average American student according to data from the Census Economic Bureau and the National Center for Economic Statistics. Additionally, Table 1 includes summary statistics for our outcome variables. On a zero to 100 scale, the students averaged an 85 on their high school GPA and 95 percent of our sample of high school seniors graduated high school. Only half of the sample, however, attempted to take the SAT but about 60 percent enrolled in college after graduation. Of these, 43 percent enrolled in a 4-year college and 40 percent did so full time.

Table 2 shows descriptive statistics for character skills measures. Concerning measures of survey effort, on average, students did not answer two percent of the items they were asked to complete. Forty-seven percent of students answered all questions in the survey. These item non-response rates are similar to those found by Hitt, Cheng, and Trivitt (2016) in multiple national representative samples of adolescents. Our careless answering measure, which captures inconsistent responses, ranges from -2.2 to 3.9. This indicates considerable variation in the degree of care that students put into completing the surveys with some being more careful than the average (negative values) and some being considerably less careful (positive values). Finally, concerning performance task measures, students devoted an average of 64 percent of the assigned time (about 10 minutes) in the diligence task, engaging in the math exercises. They spent an average of 55 percent of the assigned time (almost three minutes) tracing instead of engaging with the distractors when completing the frustration task.

With respect to self-reports and teacher-reported measures, the average self-reported grit of students in our sample was almost four. Similarly, our students scored an average of almost four on the self-control combined scale. Additionally, teachers reported an average of 3.5 in the levels of the grit of students in our sample, 3.7 in the level of work-related self-control and 4.2 in

the level of interpersonal self-control. Teachers reported that students needed redirection on average about once during the previous week and they completed, on average, about 78 percent of the assigned homework on time and with a passing grade.

#### ***4.2 Relationship among Character Skills Measures***

Table 3 presents Spearman correlations among our proposed survey effort measures and student self-reported and teacher-reported measures of character skills. As expected, item non-response rates and careless answering were negatively correlated with self-reported grit and self-control, as well as teacher reported grit, self-control, and homework completion. Additionally, both survey effort measures were positively correlated with teacher redirection. Importantly, they were both negatively correlated with performance on both the diligence and frustration tasks which corresponds with what we expected (i.e. lower levels of effort on the survey correspond with lower levels of performance in these tasks).

Table 4, shows partial rank correlations among these measures after we control for students' cognitive ability and socio-demographic information. We observe a similar pattern compared to the zero-order correlations, but partial correlations with teacher reports and performance task measures were smaller. Although the magnitudes of the correlations between survey effort and survey self-reported measures may appear small, they are at least as large as the correlations reported in prior literature validating other behavioral-task measures of conscientiousness, grit, and self-control (Duckworth and Kern, 2011; Galla et al., 2014; Meindl et al., in press).

#### ***4.3 Relationship between Survey Effort Measures and Academic Outcomes***

Next, we present evidence of criterion validity, by studying the predictive power of our proposed survey effort measures on high school and college academic outcomes. For comparison, we also study the predictive power of the performance task measures. Table 5 presents the results of linear regression models for student academic outcomes, following specification (1) described above, when different survey effort measures and performance task measures are included as explanatory variables. Regressions that use SAT scores as dependent variable are limited to only those students who attempted the SAT test. Sample sizes vary depending on the available information for each individual regression model, ranging from 392

to 458 observations and from 216 to 240 for SAT score regressions. Similarly, results for the frustration task, following Meidl et al. (in press), exclude data from students if they failed to complete a practice trial preceding the actual task, fully completed tracing the shape, experienced technical problems during the task, or were not allowed an adequate amount of time to complete the task due to data collection constraints.<sup>5</sup>

Focusing on the results for survey effort measures, presented in Table 5, we observe that a standard deviation increase in item non-response leads to an almost 0.2 standard deviations decrease in high school GPA, a 0.2 standard deviations decrease in the probability of attempting the SAT, a 0.14 decrease in SAT scores if attempted, almost 0.2 standard deviations decrease in end of the year math and reading scores, and about 0.2 standard deviations decrease in the probability of being enrolled in college one year after graduation, keeping student cognitive ability and demographic information fixed. We also estimated models that include both item non-response rates and a binary indicator for leaving any question blank. The idea of this specification is to study if both behaviors are related to academic outcomes. We find this is generally the case, as both appear to be significant predictors of these academic outcomes. Similarly, one standard deviation increase in careless answering was associated with 0.12 standard deviations decrease in high school GPA, a comparable decrease in the probability of attempting the SAT, about 0.1 standard deviations decrease on end of the year math and reading exams, and a 0.08 standard deviations decrease in the probability of being enrolled in college one year after graduation, all else equal. Finally, none of the survey effort measures was found to be related to the probability of graduating high school. This result could be due to the fact that a great majority of students in our sample (95 percent) graduated high school, limiting the available variation in this variable to find an association.

Concerning the performance tasks, we found that both performance in the academic diligence and frustration task significantly predicted high school GPA and end of the year math and reading test scores. Estimated effects were also comparable in size to those we found for survey effort measures. One standard deviation increase in performance in the diligence task is associated with 0.14 standard deviations increase in high school GPA, 0.11 standard deviations increase in SAT scores, and about 0.14 standard deviations increase in end of the year math and

---

<sup>5</sup> As a robustness check, models we also perform estimates including the full data set (i.e. n=513) and the main results were comparable to the ones presented above. Results available from the authors upon request.



reading test scores. Performance in the diligence task also significantly predicted SAT scores and college enrollment, but only marginally. Finally, performance on the frustration task significantly predicted the probability of attempting the SAT. These findings confirm the work of Galla et al., (2014) and Meindl et al. (in press) who found that performance on the academic diligence task and the frustration task predicted high school academic outcomes and college enrollment.

## **5. Conclusions**

We used data from a study of high school seniors to study the potential of survey effort measures to serve as proxy measures of character skills. Surveys often resemble routine paperwork and tasks that people have to complete in their everyday lives. In particular, for students, surveys administered in schools resemble schoolwork or homework. Therefore, we hypothesize that by measuring how much effort students put forward on surveys we can learn relevant information about their diligence and conscientiousness in academic work.

We study the potential of two parametrizations of survey effort measures that have shown promise in previous literature: item non-response and careless answering. We contribute to previous research in two ways. Firstly, we complement the work of Hitt, Trivitt and Cheng (2016), and Hitt (2015) on the validity, among adolescent students, of survey effort measures by studying their correlation with teacher reports of students' skills and with academic outcomes at the end of high school and college attendance. Secondly, we also study the relationship between survey effort measures and carefully designed performance task measures of academic diligence and frustration tolerance.

Our results show the promise of survey effort measures to be used as proxy measures of character skills related to grit and self-control. Both item non-response and careless answering showed convergent validity through negative correlations with self-reported and teacher-reported measures of grit and self-control. Correlation parameters were of similar magnitude than those found for well-designed performance task measures of diligence and frustration tolerance. Item non-response also demonstrated criterion validity through significant negative correlations with high school GPA, the probability of attempting the SAT, SAT scores when attempted, performance on end of the year math and reading tests, and the probability of being enrolled in college one year after graduation. Careless answering also showed significant correlations with

senior year GPA, attempting the SAT, end of the year math and reading test scores and college enrollment.

We note one key limitation of our study in that we only used a convenience sample of high school students in the United States. We encourage further replication work using other samples and settings to corroborate our results.

We believe that this study adds evidence to the potential of survey effort measures to provide meaningful information about students' character skills. These measures provide researchers and evaluators with a relatively easy source of information on students' character skills related to grit and self-control in a manner that is not affected by biases that can affect self-reported or teacher-reported measures, as respondents are usually unaware they are being monitored on their survey effort. In addition, they open the opportunity to future character skills research of already collected data with no direct measures of these skills (see, e.g. Cheng and Zamarro, 2018).

## References

- Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). Personality psychology and economics. *Handbook of the Economics of Education*. Vol. 4, 1-181.
- Cheng, A., & Zamarro, G., (2018). Measuring teacher non-cognitive skills and its impact on students: Insight from the Measures of Effective Teaching Longitudinal Database. *Economics of Education Review*, 64, 251-260.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 174-203.
- Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45, 259-268.
- 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., & Yeager, D., (2015). Measurement matters assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44 (4), 237-251.
- Eid, M. E., & Diener, E. E. (2006). *Handbook of Multimethod Measurement in Psychology*. American Psychological Association.
- Galla, B. M., Plummer, B. D., White, R. E., Meketon, D., D’Mello, S. K., & Duckworth, A. L. (2014). The Academic Diligence Task (ADT): Assessing individual differences in effort on tedious but important schoolwork. *Contemporary Educational Psychology*, 39(4), 314-325.
- 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.
- Hedengren, D., & Stratman, T. (2012). The dog that didn’t bark: What item nonresponse shows about cognitive and noncognitive ability. Unpublished Manuscript. Retrieved from <http://ssrn.com/abstract=2194373>
- Hitt, C. (2015). Just filling in the bubbles: Using careless answer patterns on surveys as a proxy measure of noncognitive skills. EDRE working paper 2013- 05. Fayetteville, AR: Department of Education Reform, University of Arkansas.
- Hitt, C., Trivitt, J., & Cheng, A. (2016). When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*, 52, 105-119.
- Huang, J., Curran, P., Keeney, J., Poposki, E., & DeShon, R., (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39, 103–129.
- Kaufman, A. S., & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test. John Wiley and Sons, Inc.
- Marcus, B. & Schütz, A. (2005), Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and Observer Ratings. *Journal of Personality*, 73, 959–984.
- Meade, A., & Craig, S., (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–55.

- Meindl, P., Yu, A., Galla, B., Quirk, A., Haecck, C., Goyer, P., Lejuez, C., D’Mello, S., & Duckworth, A. (in press). A brief behavioral measure of frustration tolerance predicts academic achievement immediately and two years later. *Emotion*.
- Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. *Developmental Science*, 16(6), 879-893.
- 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.
- Zamarro, G., Cheng, A., Shakeel, D., & Hitt, C (2018). Comparing and validating measures of non-cognitive traits: Performance task measures and self-reports from a nationally representative internet panel. *Journal of Behavioral and Experimental Economics*, 72, 51-60.

Figure 1. Distribution of Item Non-Response Rates

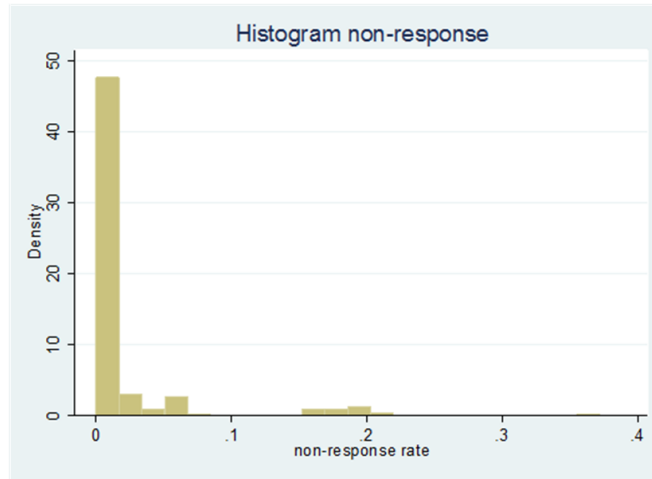
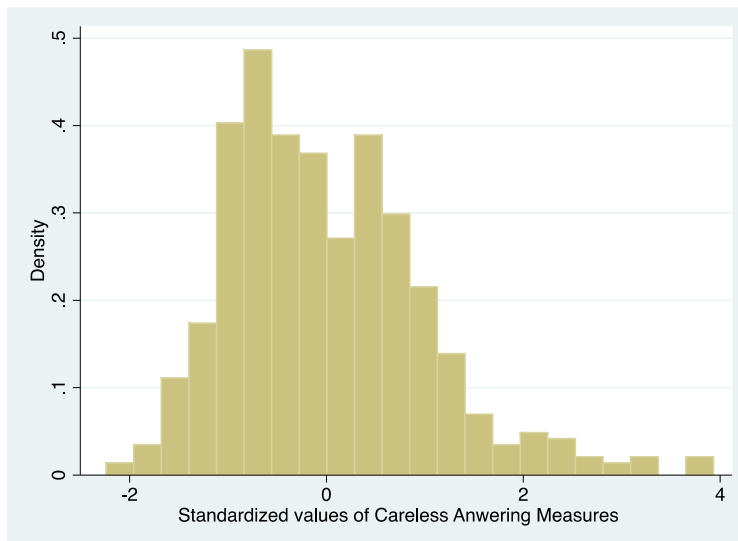


Figure 2. Distribution of Careless Answering Measures



**Table 1: Summary Statistics for Demographic and Outcome Variables**

Measure	Mean	Standard Deviation	Minimum	Maximum
<b><i>Demographic</i></b>				
Age	17.93	0.53	16	21
Female	0.54	0.50	0	1
Asian	0.20	0.40	0	1
African American	0.41	0.49	0	1
Hispanic	0.03	0.16	0	1
Caucasian	0.36	0.48	0	1
ELL	0.14	0.35	0	1
SPED	0.14	0.35	0	1
FRL	0.51	0.50	0	1
Median Household Income (\$)	52,530	22,915	9,471	128,618
KBIT Scaled Score	94.26	21.43	40	132
<b><i>Outcome</i></b>				
HS GPA Senior	85.07	7.66	55	100
HS Graduate	0.95	0.22	0	1
End of Year Math Test	1529.32	54.49	1363	1698
End of Year Reading Test	1528.77	48.32	1385	1706
Attempted SAT	0.51	0.50	0	1
Mean SAT	1414.80	254.25	820	2060
College Enrollment for 1 Year	0.64	0.48	0	1
4-year College Enrollment for 1 Year	0.43	0.50	0	1
4-year College Enrollment for 1 Year (Full-Time)	0.40	0.49	0	1

**Table 2. Summary Statistics for Measures of Character Skills**

Measure	Mean	Standard Deviation	Minimum	Maximum
<b><i>Survey Effort</i></b>				
Item Non-response	2.41	5.35	0.00	37.18
Dichotomous Item Non-response	0.53	0.50	0.00	1.00
Careless Answering	0.00	1.00	-2.24	3.93
<b><i>Performance Task Measures</i></b>				
Diligence Task Percentage Time Math	0.64	0.30	0.00	1.00
Frustration Task Percentage Time Tracing	0.55	0.27	0.00	1.00
<b><i>Self-Reported Measures</i></b>				
Grit	3.76	0.71	1.00	5.00
Locus of Control	4.57	0.75	2.50	6.00
Self-Control Combined (Work and Interpersonal)	3.61	0.60	1.00	5.00
<b><i>Teachers-Reported Measures</i></b>				
Teacher-Reported Work Self-Control	3.72	0.88	1.00	5.00
Teacher-Reported Interpersonal Self-Control	4.21	0.77	1.00	5.00
Teacher-Reported Grit	3.53	0.87	1.00	5.00
Teacher-Reported Redirection	0.92	1.16	0.00	5.00
Teacher-Reported Homework Completion	77.69	21.63	0.00	100.00

Note: Total sample size of 513 students. The statistics reported for the Frustration Task are from an analytical sample of 391 students. Following Meidl et al. (in press), we removed participants if they failed to complete a practice trial preceding the actual task, fully completed tracing the shape, experienced technical problems within the task, or were not allowed an adequate amount of time to complete the task.

**Table 3. Spearman Correlations between Survey Effort, Self-reports, Teacher Reports and Performance Task Measures of Diligence and Frustration Tolerance**

	<b>Item Non-response (1)</b>	<b>Careless Answering (2)</b>
<i><b>Self-Reported Measures</b></i>		
Grit	-0.118*	-0.024
Locus of Control	-0.093*	0.038
Self-Control Combined	-0.135*	-0.104*
Self-Control Work	-0.081*	-0.127*
Self-Control Interpersonal	-0.144*	-0.042
<i><b>Teachers-Reported Measures</b></i>		
Teacher-Reported Grit	-0.216*	-0.170*
Teacher-Reported Work Self Control	-0.201*	-0.165*
Teacher-Reported Interpersonal Self-Control	-0.147*	-0.092*
Teacher-Reported Redirection	0.112*	0.133*
Teacher-Reported HW Completion	-0.157*	-0.105*
<i><b>Performance Task Measures</b></i>		
Diligence Task PT Math	-0.152*	-0.163*
Frustration Task PT Trace	-0.104*	-0.134*

Notes: \* represents p-value<0.05. Total sample of 513 students. The statistics reported for the Frustration Task are from a sample of 391 students.



**Table 4: Partial Rank Correlations between Performance Task Measures and Self-Reports and Teacher Reports**

	Item Non-response (1)	Careless Answering (2)
<i>Self-Reported Measures</i>		
Grit	-0.155*	-0.066
Locus of Control	-0.091	-0.022
Self-Control Combined	-0.148*	-0.114*
Self-Control Work	-0.101*	-0.153*
Self-Control Interpersonal	-0.155*	-0.035
<i>Teachers-Reported Measures</i>		
Teacher-Reported Grit	-0.184*	-0.131*
Teacher-Reported work Self-Control	-0.164*	-0.107*
Teacher-Reported Interpersonal Self-Control	-0.122*	-0.065
Teacher-Reported Redirection	0.111*	0.091
Teacher-Reported HW Completion	-0.122*	-0.058
<i>Performance Task Measures</i>		
Diligence Task PT Math	-0.084	-0.125*
Frustration Task PT Trace	-0.067	-0.102*

Notes: Controls include KBIT Scaled Score, Age, Ethnicity, Gender, FRL, SPED, ELL, and household income. \* represents p-value<0.05. Total sample of 513 students. The results reported for the Frustration Task are from a sample of 391 students.

**Table 5. Standardized Coefficients of Linear Regression Models Predicting Academic Outcomes**

	High School GPA	High School Graduation	Attempt SAT	SAT	End of Year Math	End of Year Read	College Enroll 1 year	4yr College Enroll 1 year	4yr College Enroll Full Time 1 year
<b>Item Non-Response (%)</b>	-0.196*** (0.043)	0.024 (0.047)	-0.236*** (0.044)	-0.139** (0.057)	-0.193*** (0.040)	-0.197*** (0.041)	-0.238*** (0.045)	-0.213*** (0.044)	-0.192*** (0.044)
<i>Adj R-squared</i>	0.240	0.077	0.161	0.273	0.374	0.316	0.130	0.170	0.160
<b>Dichotomous Item Non-response</b>	-0.144*** (0.043)	-0.065 (0.046)	0.271*** (0.044)	-0.125** (0.057)	-0.164*** (0.039)	-0.228*** (0.041)	-0.249*** (0.045)	-0.242*** (0.044)	-0.214*** (0.044)
<i>Adj R-squared</i>	0.224	0.080	0.178	0.270	0.365	0.329	0.132	0.180	0.168
<b>Dichotomous Item Non-response</b>	-0.078* (0.047)	-0.090* (0.051)	-0.209*** (0.047)	-0.078 (0.063)	-0.103*** (0.043)	-0.176*** (0.045)	-0.183*** (0.049)	-0.186*** (0.047)	-0.164*** (0.048)
<b>Item Non-Response (%)</b>	-0.163*** (0.047)	0.061 (0.051)	-0.151*** (0.048)	-0.104 (0.064)	-0.150*** (0.043)	-0.123*** (0.045)	-0.163*** (0.049)	-0.137** (0.048)	-0.125** (0.048)
<i>Adj R-squared</i>	0.243	0.081	0.194	0.275	0.381	0.340	0.151	0.193	0.179
<b>Careless Answering</b>	-0.119*** (0.044)	-0.021 (0.047)	-0.122*** (0.046)	0.033 (0.059)	-0.133*** (0.040)	-0.097** (0.043)	-0.083* (0.047)	-0.058 (0.046)	-0.047 (0.046)
<i>Adj R-squared</i>	0.217	0.076	0.122	0.255	0.355	0.287	0.079	0.127	0.126
<b>Diligence Task PT Math</b>	0.145*** (0.047)	-0.021 (0.051)	0.023 (0.050)	0.114* (0.063)	0.144*** (0.044)	0.126*** (0.046)	0.093* (0.051)	0.064 (0.049)	0.057 (0.05)
<i>Adj R-squared</i>	0.232	0.077	0.110	0.242	0.352	0.304	0.091	0.147	0.127
<b>Frustration Task PT Tracing</b>	0.132*** (0.050)	0.052 (0.053)	0.118** (0.053)	0.046 (0.069)	0.196*** (0.049)	0.117** (0.051)	0.086 (0.054)	0.065 (0.053)	0.054 (0.053)
<i>Adj R-squared</i>	0.208	0.114	0.117	0.171	0.278	0.223	0.068	0.095	0.100

Note: Standard errors in parenthesis. Additional controls included in the model are: KBIT Scaled Score, Age, Ethnicity, Gender, FRL, SPED, ELL and household income; \* indicates P-values<0.1, \*\* indicates P-values<0.05, and \*\*\* indicates P-values<0.01.