Are Students On Track?: Comparing the Predictive Validity of Administrative and Survey Measures of Cognitive and Noncognitive Skills for Long-Term Outcomes

Christopher Cleveland*
Brown University
christopher_cleveland1@brown.edu

Ethan Scherer
Center for Education Policy Research and
READS Lab at Harvard University
ethan_scherer@gse.harvard.edu

January 2025

Abstract

Education leaders need valid metrics to predict students' long-term success. We use a unique dataset with cognitive skills, self-regulation, behavior, course performance, and test scores for 8th-grade students from a Northeast school district. We link these data to students' high school outcomes, college enrollment, persistence, and on-time degree completion. Survey-based cognitive and self-regulation measures predict high school and college outcomes. However, these relationships become small and lose statistical significance when test scores, GPA, and an absences-suspensions index are included in the predictive models. For leaders hoping to identify the best on-track indicators for college completion, the information collected in student longitudinal data systems better predicts both short- and long-term educational outcomes than the survey-based self-regulation and cognitive measures.

* Indicates correspondence author

Acknowledgements:

Both authors contributed equally to this manuscript. We thank Martin West for project support and guidance, Rebecca Ivester for assistance in this research, the four anonymous reviewers for thoughtful comments, and conference participants at the APPAM and AEFP conferences. We thank the Massachusetts Department of Elementary and Secondary Education and Charter Management Organizations for providing the administrative data for this work. We thank the Walton Family Foundation and the Bill and Melinda Gates Foundation for financial support. In addition, the research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150010 to Harvard University. The opinions expressed are those of the authors and do not represent the views of the Institute, the U.S. Department of Education, the Walton Family Foundation, or the Bill & Melinda Gates Foundation. The data used in the study is confidential and must be requested from the Center for Education Policy Research at Harvard University, Massachusetts Department of Elementary and Secondary Education, and Charter Management Organizations. Any errors of fact or interpretation are our own.

Introduction

Education leaders have traditionally focused on test scores to inform decisions about how to support students. Standardized tests predict later outcomes, but there is a growing belief that they do not provide the whole picture (Kuncel & Hezlett, 2007; Sackett et al., 2008; Goldhaber et al., 2020). Studies document that high-school GPA outperforms SAT/ACT scores in predicting college graduation, even without adjusting for high-school quality (Geiser & Santelices, 2007; Bowen et al., 2009; Galla et al., 2019). Due to the subjectivity of grades across schools, these results are surprising (Camara, 2003). However, research suggests that noncognitive competencies (i.e., socioemotional or social-emotional), like self-regulation, are differentially crucial for earning course grades compared to standardized tests (Duckworth et al., 2012; Liu et al., 2023). Farrington et al. (2012) state that conscientiousness, self-control, and grit are measures of academic perseverance. These factors lead students to engage in pro-academic behaviors, such as attending school and studying. These competencies provide incremental predictive power for essential metrics like on-time graduation (Galla et al., 2019). This evidence broadens our understanding of ensuring future success and has pushed policymakers to find additional measures to support students.

The field has proposed several ways to capture these noncognitive competencies. Many districts collect student, teacher, or parent reports of these competencies. However, there are well-documented concerns surrounding self-reported measures: (1) social desirability bias because students want to be viewed favorably by their teachers (Duckworth & Yeager, 2015), (2) outsiders' ratings are generally more predictive of individuals' behaviors than individuals' ratings (Connelly & Ones, 2010; Oh et al., 2011; Poropat, 2014), and (3) reference bias – the rating is relative to your peers rather than absolute (Duckworth & Yeager, 2015; West et al., 2016). These concerns have led researchers to look at administrative data (Kautz & Zanoni, 2014; Heckman et al., 2018; Jackson, 2018). Kautz and Zanoni (2014) create a noncognitive proxy that factors students' grades, credits, suspension, expulsions, and absences. Liu et al. (2023) use data from the CORE districts, a consortium of California districts that report multiple academic and socioemotional learning indicators, and find that academic behaviors, like absences and suspensions, are more predictive than socioemotional competencies for various outcomes, including high school graduation and college enrollment. However, survey and administrative measures might capture different underlying constructs due to when they are captured. For example, GPA captures the culmination of a year's worth of academic knowledge, including turning in homework and studying for tests. A student might have started the year failing to turn in assignments and receiving failing grades. Failing signals the students that they need to improve. If these signals motivate change, the student's self-reported self-regulation could change over the year. Capturing survey measures at specific ages could yield different outcomes as well. Using a sample of 282,867 CORE Students in California, West et al. (2020) find there are normative changes in students' self-reports of the competencies, particularly during the middle school grades, tied to academic and behavioral outcomes (Kanopka et al., 2024). Given the potential malleability of the survey and behavioral measures, it remains unclear whether a measurement at a point in time could be meaningful for long-run outcomes. For example, a student might answer a survey question differently on Mondays versus Fridays, depending on how they experienced the week, or answer differently in the Fall when they are excited about school and seeing their friend compared to late Spring when they might be thinking more about summer. To our knowledge, no study has computed self-regulation using administrative data and compared it to the self-reported survey measures to assess its predictive validity for college graduation.

Relatedly, there is concern that the predictive validity of test scores on post-secondary success may only reflect differences in students' underlying cognitive abilities rather than something more malleable. Cognitive skills like processing speed (how fast to carry out simple cognitive tasks), working memory (the amount of information that can be processed and kept in mind), and fluid reasoning (the flexibility to problem solve in a novel domain) tend to be rank-stable after age 10 (Heckman & Mosso, 2014). If test scores only measure cognitive ability, schools might be able to do little to improve these scores after age 10. Goldhaber et al. (2020) find that 3rd-grade test scores predict high school outcomes with a high degree of accuracy, suggesting that perhaps academic achievement is like cognitive skills. However, Finn et al. (2014) find evidence that standardized tests measure something different than cognitive skills. They show that schools improved math achievement but did not impact students' cognitive abilities, suggesting that test scores might be more malleable than cognitive measures assessed after 4th grade. Other researchers have also found that student achievement tests are sensitive to classroom teachers and other factors (Chetty et al., 2014; Jacob & Rothstein, 2016; Soland et al., 2019; Ang, 2020). Because it is rare to have standardized tests, cognitive skills, and college outcomes in the same dataset, whether achievement test scores predict college success only through cognitive skills remains an undeveloped area of research.

We address these unanswered questions by using a unique dataset of 1,340 8th-grade students attending a Northeast district containing administrative data on behavior and course performance, survey-based measures of self-regulation (i.e., conscientiousness, grit, and self-control), and measures of cognitive skills (both test scores and measures of processing speed, working memory, and fluid reasoning). We chose eighth grade because it precedes the critical transition to high school and is one of the lowest points for student self-reports on self-regulation (West et al., 2020). We aim to understand better how these metrics relate to high school and college outcomes. Specifically, some measures might be more strongly associated with some outcomes. We also hypothesize that the timing of the measures matters. Administrative measures are collected over a year or represent accumulated knowledge over a year. These measures can capture feedback students receive (e.g., I failed a test, a notice was sent home to my parents for being chronically absent) and then average the student's changes over a year. Survey measures are taken during a particular time – a snapshot of how a student feels that day. Because all these measures are in one dataset, it provides a complete picture of how they interact. This study can inform education leaders on how measures in a critical grade (8th) predict longer-term outcomes.

2. Data and Measures

Administrative Data

We collected data from 8th-grade students attending a Northeast district's middle schools during the spring semester of the 2010-2011 school year. Within these schools, we include all students from whom we received parental consent to participate and who attended school on the data collection day. Students completed cognitive tests and surveys assessing their self-regulation abilities in their regular classrooms. We follow this cohort of students through the 2018-19 school year to measure our longest-term outcome of on-time BA degree completion.

We merged these data with student-level high school administrative data—enrollment, attendance, suspensions, grade point average (GPA), math and English language arts (ELA) state test scores, and typical demographic information. For the whole district, the scaled scores were standardized by grade, subject, and year to have mean zero and variance one. Of the 3,723 8th-grade students in the district with complete administrative and demographic data, 2,586 attended a surveyed school.[1] The analytic sample is restricted to the 1,340 survey participants with complete administrative and demographic information. Of the 2,586 students in the sampled schools, approximately 68% of the parents consented to be in the study, and of these students, 76% completed both the cognitive and noncognitive surveys.

We also received student-level college information from the National Student Clearinghouse (NSC) for 2012 to 2019, so the college outcomes are restricted to within four years of a student's expected high school graduation. College-going is defined as enrolling in a two- or four-year college, and college persistence is the number of college quarters attended. A quarter is a student having an enrollment start or end date in one of four three-month periods—i.e., January to March. When defining quarters, end dates supersede start dates. We define bachelor's degree completion as a student graduating from a four-year college.

The demographic characteristics of the sample are in Table 1. Table 1 compares demographic and 8th-grade behaviors and outcomes of students between all 8th-grade students to the subset of students who attended a school that administered surveys and to the specific subset of students who consented. Our final analytic sample is racially and ethnically diverse; 71% are Hispanic or Black and come from low-income families, with 81% of students receiving free or reduced-price lunch. These participants differ from the district sample. The surveyed students outscored the district schools by 0.29 standard deviations in math, 0.25 standard deviations in ELA, and 0.19 in GPA. They had fewer absences by 3.22 days and 0.07 fewer suspensions. These differences suggest a positive selection into the survey sample. This selection into survey

[1] We leveraged data from a prior study using charter lottery data. The study included any school students attended (both students with and without lottery offers). Table 1 shows these schools tended to be higher performers.

participation may limit the generalizability of our findings as parents who consent and students who participate in the survey might have families more invested in school. These unobserved characteristics could lead to stronger relations with academic and educational attainment outcomes than those families that did not consent.

Absences-Suspensions Index & GPA

Our work compares survey and test-based measures with behavioral data already collected regularly by school systems. Following (Kautz & Zanoni, 2014; Heckman et al., 2018; Jackson, 2018), we proxy for socioemotional competencies by using non-test score behavior available in the state data: the log of the number of absences in 8th-grade, whether the student was suspended in 8th-grade and 8th-grade GPA. These behaviors are associated with commonly used self-regulation measures at scale (West et al., 2020). Then, like Jackson (2018), we use principal component analysis (PCA) to create an index but separate grade point average from the other measures because it represents both cognitive (e.g., performance on a test) and behavior (e.g., did the student study for the test). We add one to each student's absence count to avoid cases where a student has no absences, and the natural log is undefined. The PCA retained a single factor (eigenvalue = 1.21). The factor loadings for each variable are as follows: natural log of absences = 0.78 and ever suspended = 0.78. We then predict scores for each student using the Bartlett method and standardize the final index to have a mean zero and variance one (Jackson, 2018). The absences-suspensions index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes.

Self-Regulation Skills Measures

The study includes three self-regulation skills measures: conscientiousness, self-control, and grit. We use these measures to extend Farrington et al. (2012) and Finn et al. (2014), highlighting how these measures are important for elementary and secondary academic outcomes.

Conscientiousness

To evaluate conscientiousness, students completed the Big Five Inventory, which assesses five personality traits—neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (John & Srivastava, 1999). Students rate how much they agree or disagree with a statement (1 = strongly disagree to 5 = strongly agree). Nine items relate to conscientiousness (e.g., "I think I am someone who is a reliable worker"). A student's conscientiousness score is calculated by taking the average of their ratings on these items. This scale had an internal reliability of 0.75.

Self-Control

To assess self-control, students completed the Impulsivity Scale for Children (Tsukayama et al., 2013), which has eight items that measure students' impulsivity as it relates to behavior, attention, and emotions. Items asked students to indicate how often, on a five-point scale ranging from "almost never" to "at least once a day," they exhibited specific behaviors in the past year. Four items assessed interpersonal self-control (e.g., "I interrupted other students while they were talking"), and four assessed intrapersonal self-control (e.g., "I forgot something I needed for class"). We calculated their overall impulsivity score by reverse-coding and averaging these eight items. This scale had an internal reliability of 0.83.

Grit

Grit is "perseverance and passion for long-term goals" (Duckworth et al., 2007) using the Short Grit Scale (Duckworth & Quinn, 2009). Students respond to eight items (e.g., "I finish whatever I begin") on a five-point scale that ranges from "not like me at all" to "very much like me." Overall grit is calculated by taking the average score for these eight items. This scale had an internal reliability of 0.64.

We then use PCA for the three self-regulation skills and refer to the composite of the three skills as our 'self-regulation index.' The PCA retained a single factor (eigenvalue = 2.06). The factor loadings for the three measured self-regulation abilities include conscientiousness = 0.61, impulse control = 0.52, and grit = 0.59. We then predicted scores using the Bartlett method for each student and standardized them to have mean zero and variance one (Jackson, 2018).

Cognitive Skills Measures

The study includes three cognitive skills measures: processing speed, working memory, and fluid reasoning. These measures are derived from previous studies regarding the cognitive skills of school-age students, which are positively correlated with standardized test scores (Finn et al., 2014).

Processing Speed

We use the Coding and Symbol Search subtests from the fourth edition of the Wechsler Intelligence Scale for Children to evaluate processing speed (Wechsler, 2003). For the coding task, students were given a key that assigned the digits 1 through 9 unique symbols and asked to translate a string of digits to the corresponding symbols. For the symbol search task, students were asked to determine whether the two symbols on the left side of the page matched any of the five symbols on the right. Students had two minutes for each task.

Working Memory

We use the count span task to measure working memory (Case, Kurland, & Goldberg, 1982; Cowan et al., 2005). For this task, students were presented with a display of blue circles, blue triangles, and red circles and were instructed to note the number of blue circles. Students viewed as few as one and as many as six such displays before being asked to recall the number of blue circles in each display of the series. The task begins with a single display and increments by one for every three consecutive trials of a given load the student gets correct and maxes out at a load of six. Students were given 4.5 seconds to note the number of blue circles in each display.

Fluid Reasoning

We use the fourth edition of the Test of Nonverbal Intelligence to measure fluid reasoning (Version A; Brown et al., 2010). Students chose which of the six pictures completed the given puzzle. Students were given ten minutes to complete as many as 40 puzzles. Puzzles increased in difficulty as the students progressed through the task.

Due to the high correlation between these three measures, we use PCA to create a composite of these measures that we refer to as the 'cognitive index.' The PCA retained a single factor (eigenvalue = 1.60). The factor loadings for the three measured cognitive abilities include processing speed = 0.61, working memory = 0.53, and fluid intelligence = 0.60. We then predicted scores using the Bartlett method for each student and standardized them to have mean zero and variance one (Jackson, 2018).

3. Empirical Strategy

To evaluate whether our measures collected in the Spring of 8th-grade are predictive of high school and college outcomes, we estimate:

$$Y_{i,t+n} = \beta_1 SR_{it} + \beta_2 C_{it} + \beta_3 AS_{it} + \beta_4 GPA_{it} + \beta_5 MATH_{it} + \beta_6 ELA_{it} + \delta X_{it} + \sigma_s + \epsilon_{it}$$

$Y_{i,t+n}$ represents student *i*'s outcome in a period after period t. $SR_{it}\ and\ C_{it}$ represent student i's self-regulation index and cognitive index scores, respectively, at time t in 8th grad$e,$ AS_{it} represents the student's absences – suspensions index score, GPA_{it} represents the student’s GPA and $MATH_{it}$ and ELA_{it} the student's math and ELA standardized scores. X_i

represents a vector of a student's characteristics, including gender, race, economic disadvantaged status, special education status, and English language learner status. σ_s represents school fixed effects. Standard errors are clustered at the 8th-grade school. Because GPA practices vary across schools (Camara et al., 2003) and prior research has found that GPA predicts on-time college graduation (Galla et al., 2019), we ran our models with and without the school fixed effects. We do not find a substantial difference between these two models and present the school fixed effects model.

4. Results

Correlations of Measures

Table 2 shows Pearson correlations among the cognitive measures, self-regulation measures, three composite measures, GPA, and 8th-grade ELA and math test scores.

While each cognitive test measures a different aspect of cognitive skills, we see significant correlations (0.27 to 0.36) among them. The Fry and Hale (1996) developmental cascade theory suggests that changes in processing speed are related to changes in working memory that drive changes in fluid intelligence, which makes the weak to moderate relationships among these skills reasonable. Our three self-regulation measures show moderate correlations (0.42 to 0.67) among the competencies that reflect academic perseverance.

The absences-suspension index is most strongly correlated with self-regulation (0.21) since the components of the behavioral index—absences and suspensions—largely reflect students' self-regulation competency. GPA correlates most strongly with the self-regulation and absences-suspensions index, demonstrating its association with noncognitive competencies, but it also correlates with 8^{th}-grade math test scores (0.36).

The 8th-grade test score measures also exhibit significant correlations with other measures. Math scores correlate most with fluid reasoning (0.55) and the overall cognitive index (0.59). ELA scores correlate most with fluid reasoning and processing speed (0.39) and the overall cognitive index (0.43). The Math and ELA test scores have the lowest correlation with conscientiousness and grit, (0.03) and (0.01), respectively.

In sum, cognitive measures correlate strongest with each other, as do self-regulation measures that reflect academic perseverance. The absences-suspensions index correlates strongly with self-regulation, providing preliminary evidence that administrative data may be a reasonable proxy for socioemotional competencies.

High School Outcomes

We first examine how our administrative data and cognitive and self-regulation (SR) indices predict 10th-grade math and ELA test scores (Table 3, models 1-5 and 6-10, respectively). For each outcome, the first model only includes administrative data. The next model only includes the cognitive and self-regulation survey measures. In the third model, we add the absences-suspensions index and GPA with the cognitive and self-regulation index to assess whether the administrative proxy and self-reported self-regulation competencies independently predict the outcomes, controlling for cognitive skills. Next is a model with 8th-grade test scores and cognitive and self-regulation survey measures. The final model includes all measures.

In Table 3, models (1), (2), (6), and (7) demonstrate that when either the survey measures or the administrative measures are added to the model, both the cognitive and noncognitive measures independently predict 10th-grade standardized scores in math and ELA.[2] In both cases, the magnitude of the effect of the “noncognitive” component of the regression is much smaller than the cognitive component but still independently predicts future test score performance. For example, GPA is independently predictive of test scores even when controlling for prior test scores in models (1) and (6). However, only the administrative construct remains significant at the 0.05 level when the administrative measures are compared to the survey-based ones. While we expect 8th-grade scores to be strongly correlated with 10th-grade scores, the fact that the cognitive skills offer no additional information is interesting because test scores are often considered synonymous with cognitive skills. We observe that other measures, like GPA, provide little additional information to the prediction.

We next examine the high school dropout and graduation outcomes (Table 4, models 1-5 and 6-10, respectively). In models (1), (2), (6), and (7), only the noncognitive measures independently predict dropout and graduation. However, when we include both the survey data and the absences-suspensions index and GPA, models (3) and (8), respectively, the self-regulation index loses significance, suggesting the administrative data are better proxies for the competencies than the survey measures.

In model (9), we remove and replace the absences-suspensions index and GPA with test scores. Math test scores, but not ELA test scores, explain additional variance in high school graduation but not dropout. Passing the 10th-grade standardized test is required to graduate high school in the state. However, this requirement is for both Math and ELA test scores, and the

[2] Note that not all students have MCAS scores because they could dropout or be absent on the test date, but they will have graduation and dropout data.

relationship between ELA scores and graduation is significant at the 0.10 level. Thus, the Math test scores could proxy for cognitive and noncognitive skills other than self-regulation because the cognitive and self-regulation indices remain statistically significant, and the magnitude is unchanged. As observed in regression (6), there is no independent predictive power of math test scores once the behavioral index is included. Graduating from high school involves getting homework in on time and attending classes, which requires self-regulation, among other competencies. Thus, its dominance over all the other measures seems reasonable.

College Outcomes

We next examine two- or four-year college enrollment, college persistence, and bachelor's degree completion (Table 5, models 1-5, 6-10, and 11-15, respectively). In models (1), (2), (6), and (7), we observe that the noncognitive competencies most strongly predict college enrollment and persistence, which requires students to submit the appropriate forms on time, attend classes, and manage their newfound freedom or constraints among other logistical hurdles. What is surprising is that 8th-grade ELA is independently predictive of college enrollment and persistence. College essays and enrollment require correspondence, and most majors require some writing; however, why ELA is predictive but not mathematics is unexpected. Further, as we observed with graduation and dropout, once the absences-suspensions index and GPA are added to the regression, the self-regulation index is no longer independently predictive for enrollment or persistence in college at the 0.05 level. These findings align with what Liu et al. (2023) found.

Only GPA is significant for predicting bachelor's degree completion in model (11). The cognitive and self-regulation indices are significant predictors in model (12). When we have all the predictors in the model, GPA is the only significant predictor (15). The fact that GPA from eight years prior can predict a nine-percentage point increase in on-time bachelor's degrees demonstrates the power of the tools that schools already have. The absences-suspensions index contributes to predictive power but is only significant at the 0.10 level.

In all models for all outcomes, survey measures alone explain the least variation and have the worst fit based upon adjusted R^2, Akaike information criterion (AIC), and Bayesian information criterion (BIC). Other than math, models with administrative data alone explain the most variation and have the best fit.

Additional Analyses

We provide additional analyses in the Online Appendix to assess the primary results and sample. In Online Appendix Table A1, we replicate our college outcomes analyses using administrative measures for all 8th-graders in the district. The patterns in our primary results remain unchanged with this larger sample. In Online Appendix Table A3, we replicate the high school graduation, high school dropout, and college enrollment analysis without cognitive measures and use self-management measures instead of the self-regulation index using a sample of 2015 and 2016 8th-grade students. This second sample is described in Online Appendix Table A2. In Online Appendix Table A4, we replicate the high school graduation and college enrollment analysis and explore the other types of noncognitive/socioemotional measures added to the model with the second sample. These analyses reaffirm the importance of self-regulation for academic perseverance in the outcomes we can measure through college enrollment. Item response theory (IRT) graded response model (GRM) scores were also developed for the self-regulation surveys for the primary sample. College outcomes results for the primary sample using GRM scores are in Online Appendix Table A5. The correlation between the GRM and

PCA indices is 0.97. The results do not differ between the two approaches, so we report the PCA-based index results as the primary results to be comparable to prior research in this area.

5. Discussion and Conclusion

Our work builds on the evidence that cognitive and noncognitive competencies predict future educational attainment. We replicate research that finds noncognitive measures outperform cognitive assessments in predicting degree completion (Geiser & Santelices, 2007; Bowen et al., 2009; Jackson, 2018; Galla et al., 2019). A standard deviation increase in the absences-suspensions index is associated with a three-percentage point increase in on-time BA completion at the 0.10 significance level. GPA predicts a nine-percentage point increase in on-time bachelor's degree completion. In contrast, 8th-grade mathematics and ELA test scores do not independently predict degree completion. However, these results highlight the importance of multiple measures of development. For example, with GPA included in the predictive regression, neither math nor ELA scores add independent information. However, if GPA is excluded, a one standard deviation increase in 8th-grade math scores improved on-time BA completion by eight percentage points. Similarly, when the self-regulation index is included with the cognitive index, it predicts enrollment, quarters, and BA completion. Thus, these predictive models have a real risk that omitted factors could lead to different conclusions.

Second, although students' feelings about these competencies can grow and decline over time based on the feedback they receive and that students gain knowledge in elementary and secondary school, we replicate and extend findings that point-in-time measures can predict later outcomes. Our assessment spans early secondary through a bachelor's degree and finds that measures collected early can be highly predictive of high school test score performance, dropout and graduation, and post-secondary outcomes. Given prior work showing that particularly noncognitive competencies, like self-regulation, are malleable to feedback during secondary schooling, these results are surprising but demonstrate that practitioners have a powerful tool early in a student's secondary education to develop on-track indicators commonly used in many districts and states (Easton et al., 2007; Goldhaber et al., 2020; Liu et al., 2023; Canbolat, 2024).

Furthermore, by comparing the administrative and survey-based measures in the cognitive and self-regulation domains, we can assess which measures are most predictive for different outcomes. For both high school and college outcomes, once we add the 8th-grade absences-suspensions index, GPA, and test scores to our prediction regressions with the cognitive and self-regulation indices, the self-regulation measures lose their statistical significance, suggesting they do not add explanatory power over what is available in the 8th-grade administrative data. Other recent work has emphasized the importance of precisely leveraging the state longitudinal systems for this purpose (Austin et al., 2020; Goldhaber et al., 2020). These results also help researchers and policymakers because they can leverage existing administrative measures to evaluate prior interventions in contexts where these relatively new measures, like self-assessed self-regulation, are unavailable and still gather meaningful insights. For agencies, both time and money are scarce resources; thus, diverting staff time to compile and better utilize existing data could be beneficial.

We also find evidence that absences and suspensions predict long-run outcomes, even while excluding GPA. The absences-suspensions index remains independently predictive of post-secondary enrollment and persistence. Thus, even though most prior work has focused on GPA as a predictor of college success, our evidence indicates that adding other administrative measures like absences and suspensions can strengthen this measure because they independently predict post-secondary outcomes (Geiser & Santelices, 2007; Bowen et al., 2009; Galla et al., 2019).

We also find evidence that test scores operate through more than cognitive skills. In regressions where we control for the self-regulation index, cognitive index, and 8th-grade test scores, the cognitive index is no longer significant, suggesting it does not add explanatory power over what is available in the 8th test scores. Standardized tests are not only measures of a student's aptitude. The tests capture a student's accumulated knowledge based on state content standards in core academic subjects from educational experiences, teaching, and aptitude. Furthermore, there is evidence that metadata captured by standardized tests can also capture noncognitive skills like self-efficacy, self-regulation, conscientiousness, and grit (Soland et al., 2019). We observe evidence of this finding in our data. While standardized tests are correlated most strongly with processing speed and fluid reasoning, they correlate with self-control. The fact that test scores are independently predictive of post-secondary success, even controlling for the cognitive index, is comforting, given prior research that schools might have limited ability to improve cognitive skills but might be able to improve standardized tests (Finn et al., 2014).

Because our dataset contains these four measures, we can also begin to unpack what competencies are captured in administrative data. Several recent studies have demonstrated the association between self-regulation and grade point average (Duckworth et al., 2012; Galla et al., 2019; West et al., 2020). We observe a similar pattern for the 8th-grade grade point average. However, GPA also correlates with our cognitive index, combining working memory, processing speed, and fluid reasoning. In contrast, the absences-suspensions index correlated with our three socioemotional measures (0.11-0.21) and processing speed, working memory, and fluid reasoning (0.11-.016). However, when we assess how the self-regulation index correlates with the cognitive measures, these are lower, and only working memory is statistically significant. In addition, GPA has a higher correlation with the socioemotional (0.29-0.36) and cognitive measures (0.13-0.24). If policymakers hope to use the administrative data to inform education decisions using the administrative data, they need to know what these measures capture more precisely. For example, in future research, it will be interesting to explore whether suspensions, absences, and GPA are more correlated with other noncognitive measures. Kanopka et al. (2024) only explore test scores and absences but find that self-efficacy and growth mindset are stronger predictors of attendance changes overall than self-management, but that changes in self-management (e.g., our self-regulation measure) are equally important for those with the lowest attendance.

If practitioners only consider creating the most predictive on-track systems, our research suggests that early secondary administrative data could be better than including the survey measures. We hypothesize that the end-of-year behavioral averages could be a more stable measure, making them better predictors of future outcomes, such as graduation. However, surveys could be better indicators for more immediate decisions. If a student takes a formative standardized test at the beginning and end of the year when they experience significant knowledge growth, the end-of-year test would be a better indicator than the average of the beginning and end of the year. Similarly, there is little information about students' noncognitive competencies at the beginning of the year. Thus, a survey administered in the Fall will provide more information than administrative data from the prior Spring. To this end, practitioners could use surveys formatively to intervene on students' trajectories versus waiting for end-of-year measures. Furthermore, test scores, GPA, and our behavioral index are blunt instruments that cannot help us understand the change mechanism. For example, consider if an agency decides to implement a new curriculum to improve teacher and student relationships, which results in a decline in absences. Knowing how to improve the program without additional data is challenging. In an era where parents are concerned about too much testing and surveying of students, it is worthwhile knowing that data already collected can be used to assess cognitive and noncognitive competencies. However, we argue that they must be complemented with survey-based measures to support students' needs best.

References

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. (2011) Personality psychology and economics. In: Hanushek EA, Machin S, Woessmann L, eds. Handbook of the Economics of Education. Amsterdam, Netherlands: Elsevier; 2011:1–181.

Desmond Ang, The Effects of Police Violence on Inner-City Students, The Quarterly Journal of Economics, Volume 136, Issue 1, February 2021, Pages 115–168, https://doi.org/10.1093/qje/qjaa027

Austin, Wes, David Figlio, Dan Goldhaber, Eric Hanushek, Tara Kilbride, Cory Koedel, Jaeseok Sean Lee, Jin Lou, Umut Özek, Eric Parsons, Steven Rivkin, Tim Sass, and Katharine Strunk (2020). Where are Initially Low-performing Students the Most Likely to Succeed? A Multi-state Analysis of Academic Mobility (Preliminary Draft). CALDER Working Paper No. 227-0220.

Bowen, William G., Matthew M. Chingos, & Michael S. McPherson (2009). Crossing the finish line: Completing college at America's public universities. Princeton University Press.

Brown, Linda, Rita J. Sherbenou, and Susan K. Johnsen, (2010). Test of Nonverbal Intelligence—Fourth edition (Version A). San Antonio, TX: The Psychological Corp.

Camara W., Kimmel E., Scheuneman J., Sawtell E. A. (2003). *Whose grades are inflated?* Retrieved from https://files.eric.ed.gov/fulltext/ED563203.pdf

Canbolat, Y. (2024). Early Warning for Whom? Regression Discontinuity Evidence From the Effect of Early Warning System on Student Absence. Educational Evaluation and Policy Analysis, 0(0). https://doi-org.ezp-prod1.hul.harvard.edu/10.3102/01623737231221503

Case, Robbie, D. Midian Kurland, and Jill Goldberg (1982). Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology, 33, 386–404.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104 (9): 2593–2632**.**

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. The Quarterly Journal of Economics. 126(4), 1593-1660.

Connelly, Brian S., & Deniz S. Ones (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. Psychological Bulletin, 136(6), 1092-1122.

Dee, Thomas S., and Martin R. West (2011). The noncognitive returns to class size. Education Evaluation and Policy Analysis, 33(1), 23-46.

Duckworth, Angela Lee, & Patrick D. Quinn (2009). Development and validation of the Short Grit Scale (Grit-S). Journal of Personality Assessment, pp. 91, 166–174.

Duckworth, Angela Lee, Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92, 1087–1101.

Duckworth, Angela Lee, Patrick D. Quinn, and Eli Tsukayama (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. Journal of educational psychology, 104(2), 439.

Duckworth, Angela Lee, and David S. Yeager (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. Educational Researcher, 44(4), 237-251.

Dweck, Carol (2006). Mindset: The new psychology of success. New York, NY: Random House.

Easton, John Q., Esperanza Johnson, and Lauren Sartain (2017). The predictive power of ninth-grade GPA. Chicago, IL: University of Chicago Consortium on School Research.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Finn, Amy S., Matthew A. Kraft, Martin R. West, Julia A. Leonard, Crystal E. Bish, Rebecca E. Martin, Margaret A. Sheridan, Christopher F. O. Gabrieli, and John D. E. Gabrieli (2014). Cognitive Skills, Student Achievement Tests, and Schools. Psychological Science, 25(3), 736–744. https://doi.org/10.1177/0956797613516008

Fry, A. F., & Hale, S. (1996). Processing Speed, Working Memory, and Fluid Intelligence: Evidence for a Developmental Cascade. *Psychological Science*, *7*(4), 237–241. http://www.jstor.org/stable/40062952

Galla, Brian. M., Elizabeth P. Shulman, Benjamin D. Plummer, Margo Gardner, Stephen J. Hutt, J. Parker Goyer, Sidney K. D'Mello, Amy S. Finn, and Angela L. Duckworth (2019). Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. American Education Research Journal, 20(10), 1-39.

Geiser, Saul, and Maria V. Santelices (2007). Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. Research & Occasional Paper Series: CSHE. 6.07. Center for studies in higher education.

Goldhaber, Dan, Malcolm Wolff, and Timothy Daly (2020). Assessing the Accuracy of Elementary School Test Scores as Predictors of Students' High School Outcomes. CALDER Working Paper No. 235-0520

Heckman, James J., Eric Humphries, and Gregory Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. Journal of Political Economy, 126(S1), S197-S246.

Heckman, James J., and Stefano Mosso (2014). The economics of human development and social mobility. Annu. Rev. Econ., 6(1), 689-733.

Jackson, Kirabo (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. Journal of Political Economy, 126, 2072-2107.

Jacob, Brian, & Jesse Rothstein. (2016). "The Measurement of Student Ability in Modern Assessment Systems." Journal of Economic Perspectives, 30 (3): 85–108.

John, Oliver P., and Sanja Srivastava (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (2nd ed., pp. 102–138). New York, NY: Guilford Press.

Kanopka, K., Claro, S., Loeb, S., West, M., & Fricke, H. (2024). Are Changes in Reported Social-Emotional Skills Just Noise? The Predictive Power of Longitudinal Differences in Self-Reports. *AERA Open*, *10*, 23328584241233277.

Kautz, T. D. & Zanoni, W. (2014). Measuring and fostering noncognitive skills in adolescence: Evidence from Chicago Public Schools and the OneGoal Program. Unpublished manuscript, Department of Economics, University of Chicago, Chicago, IL.
Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. Science.

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. Social Science Research 37 (3), 888–902.

Liu, J., Kuhfeld, M., & Lee, M. (2023). Noncognitive Factors and Student Long-Run Success: Comparing the Predictive Validity of Observable Academic Behaviors and Social-Emotional Skills. Educational Policy, 0(0). https://doi.org/10.1177/08959048231209262

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. Journal of Applied Psychology, 96(4), 762-773.

Poropat, A. E. (2014). Other-rated personality and academic performance: Evidence and implication. Learning and Individual Differences, 34, 24-32.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. American Psychologist, 63(4), 215.

Soland, J., Zamarro, G., Cheng, A., & Hitt, C. (2019). Identifying naturally occurring direct assessments of social-emotional competencies: The promise and limitations of survey and assessment disengagement metadata. Educational Researcher, 48(7), 466-478.

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. Developmental Science, 16, 879–893.

Wechsler, D. (2003). Wechsler Intelligence Scale for Children—Fourth edition. San Antonio, TX: The Psychological Corp.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F. O., Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' noncognitive skills and the impact of schooling. Education Evaluation and Policy Analysis, 38(1), 148-170.

West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. B. (2020). Trends in Student Social-Emotional Learning: Evidence From the First Large-Scale Panel Student Survey. Educational Evaluation and Policy Analysis, 42(2), 279-303.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. Journal of Educational Measurement, 39(1), 1-37.

Tables & Figures

Table 1. Mean Demographic Characteristics and Measures for 8th-grade Students Among All District Schools, Sampled Schools, and Sampled Students

	(1)	(2)	(3)
	All Students in the District	All Students in Sampled Schools	Sampled Students
Male	0.51	0.50	0.45
African American	0.37	0.35	0.35
Asian/Pacific Islander	0.10	0.09	0.10
Hispanic	0.37	0.38	0.36
White, Non-Hispanic	0.15	0.16	0.18
Other	0.02	0.02	0.01
Free/reduced-price lunch	0.84	0.83	0.81
Individualized education plan	0.22	0.21	0.18
English language learner	0.18	0.17	0.15
Absences (number)	11.13	10.46	7.91
Suspension (ever)	0.13	0.13	0.10
GPA (standardized)	0.00	0.00	0.19
Math scores (standardized)	0.00	0.10	0.29
ELA scores (standardized)	0.00	0.09	0.25
N	3723	2586	1340

Notes: All samples are restricted to students with complete demographic and academic information. Sampled schools participate in the cognitive and noncognitive surveys; sampled students have valid data on the cognitive and noncognitive surveys. 8th-grade GPA is standardized within school and across all students in the district. The test scores are standardized across all 8th-grade students in the district in 2011 to have mean zero and variance one. 8th-grade suspension is a binary for whether the student was ever suspended that year. Absences = the number of absences. GPA = grade point average. ELA = English language arts.

Table 2. Correlation Matrix of 8th-grade Measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Process Speed	1											
2. Work Memory	0.27***	1										
3. Fluid Reasoning	0.36***	0.26***	1									
4. Grit	0.01	0.04	-0.04	1								
5. Self-control	0.05*	0.07**	0.11***	0.43***	1							
6. Conscientious	0.06*	0.04	-0.02	0.67***	0.48***	1						
7. Cognitive Index	0.77***	0.67***	0.75***	-0.02	0.10***	0.02	1					
8. Self-Reg Index	0.05	0.06*	0.02	0.86***	0.75***	0.88***	0.04	1				
9. Abs-Susp Index	-0.16***	-0.11***	-0.16***	-0.11***	-0.21***	-0.11***	-0.18***	-0.17***	1			
10. GPA	0.24***	0.13***	0.16***	0.29***	0.30***	0.36***	0.23***	0.39***	-0.41***	1		
11. Math	0.46***	0.28***	0.55***	0.03	0.11***	0.06*	0.59***	0.07**	-0.30***	0.36***	1	
12. ELA	0.39***	0.18***	0.39***	0.01	0.05*	0.03	0.43***	0.04	-0.23***	0.31***	0.69***	1

Notes: All individual cognitive and noncognitive subtest scores are standardized to have mean zero and unit variance. The cognitive index is a weighted average of the three cognitive subtest scores calculated by running a principal components analysis on the three subtests. Similarly, the noncognitive index is a weighted average of students' grit, conscientiousness, and self-control subtest scores and is calculated by running a principal components analysis on these three subtests. The abs-susp index is calculated by running a principal components analysis using the natural log of absences and an indicator for whether a student was suspended in 8th-grade. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. The test scores are standardized across all students in a given grade in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table 3. Predicting 10th-grade Test Scores Using 8th-grade Self-Regulation Index, Cognitive Index, and Administrative Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	10th-grade Math Scores					10th-grade ELA Scores				
Self-Reg Index		0.05***	-0.02	<0.01	-0.01		0.06***	-0.01	0.02+	<-0.01
		(0.01)	(0.02)	(0.01)	(0.01)		(0.01)	(0.02)	(0.01)	(0.01)
Cog Index		0.22***	0.16***	0.04+	0.04+		0.12***	0.07*	<0.01	<0.01
		(0.05)	(0.04)	(0.02)	(0.02)		(0.03)	(0.03)	(0.02)	(0.02)
Abs-Susp Index	0.01		<-0.01		0.01	<-0.01		-0.02		<-0.01
	(0.02)		(0.03)		(0.02)	(0.02)		(0.02)		(0.02)
GPA	0.07*		0.33***		0.08*	0.10*		0.30***		0.10*
	(0.03)		(0.06)		(0.03)	(0.04)		(0.05)		(0.04)
Math	0.56***			0.57***	0.54***	0.11***			0.15***	0.11***
	(0.05)			(0.05)	(0.04)	(0.02)			(0.02)	(0.02)
ELA	0.18***			0.20***	0.18***	0.56***			0.58***	0.56***
	(0.02)			(0.02)	(0.02)	(0.04)			(0.04)	(0.04)
N	1062	1062	1062	1062	1062	1071	1071	1071	1071	1071
Adj. R^2	0.755	0.555	0.623	0.753	0.756	0.688	0.497	0.549	0.683	0.687
AIC	1388.56	2019.27	1843.93	1395.73	1384.77	1656.68	2167.09	2050.16	1672.69	1660.63
BIC	1453.14	2073.91	1908.51	1460.32	1459.29	1721.38	2221.83	2114.85	1737.38	1735.27

Note: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, cognitive and noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student ever received free- or reduced-price lunch, ever had an individualized education plan, and ever was an English language learner. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all students in a given grade in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table 4. Predicting High School Dropout and Graduation Using 8th-grade Self-Regulation Index, Cognitive Index, and Administrative Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	High School Dropout					High School Graduation				
Self-Reg Index		-0.02**	-0.01	-0.02**	-0.01		0.04***	0.00	0.03**	0.00
		(0.01)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
Cog Index		-0.00	0.01*	0.01	0.01+		0.01	-0.02*	-0.02	-0.02*
		(0.01)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
Abs-Susp Index	0.04**		0.04**		0.04**	-0.08***		-0.08***		-0.08***
	(0.01)		(0.01)		(0.01)	(0.01)		(0.01)		(0.01)
GPA	-0.05**		-0.05**		-0.05*	0.12***		0.12***		0.12***
	(0.02)		(0.02)		(0.02)	(0.03)		(0.03)		(0.03)
Math	0.02			-0.02+	0.01	-0.02			0.07**	-0.01
	(0.01)			(0.01)	(0.01)	(0.02)			(0.02)	(0.02)
ELA	-0.01			-0.02	-0.01	0.01			0.04+	0.01
	(0.02)			(0.01)	(0.01)	(0.02)			(0.02)	(0.02)
N	1210	1210	1210	1210	1210	1210	1210	1210	1210	1210
Adj. R^2	0.111	0.073	0.113	0.078	0.112	0.201	0.103	0.204	0.120	0.203
AIC	314.04	362.73	310.82	358.09	313.97	997.58	1135.70	993.10	1113.36	996.72
BIC	359.92	398.41	356.71	403.98	370.05	1043.46	1171.39	1038.99	1159.24	1052.80

Note: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, cognitive and noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student ever received free- or reduced-price lunch, ever had an individualized education plan, and ever was an English language learner. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all students in a given grade in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table 5. Predicting College Outcomes Using 8th-grade Self-Regulation Index, Cognitive Index, and Administrative Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	College Enrollment					College Quarters					Bachelor's Degree				
Self-Reg Index		0.03*	0.00	0.02*	0.00		0.75***	0.19+	0.62***	0.20+		0.03***	0.01	0.03**	0.01
		(0.01)	(0.01)	(0.01)	(0.01)		(0.16)	(0.11)	(0.15)	(0.11)		(0.01)	(0.01)	(0.01)	(0.01)
Cog Index		0.01	-0.01	-0.02	-0.02		0.20	-0.27	-0.26	-0.33+		0.03*	0.00	-0.00	-0.00
		(0.01)	(0.01)	(0.01)	(0.01)		(0.19)	(0.16)	(0.21)	(0.19)		(0.01)	(0.01)	(0.01)	(0.01)
Abs-Susp Index	-0.05***		-0.05***		-0.05***	-1.11***		-1.09***		-1.12***	-0.03+		-0.03+		-0.03+
	(0.01)		(0.01)		(0.01)	(0.17)		(0.17)		(0.17)	(0.01)		(0.01)		(0.01)
GPA	0.08**		0.10***		0.08***	1.86***		1.97***		1.73***	0.09***		0.10***		0.09***
	(0.02)		(0.02)		(0.02)	(0.30)		(0.34)		(0.29)	(0.02)		(0.01)		(0.02)
Math	-0.00			0.07**	0.01	-0.22			1.11**	0.01	0.03			0.08***	0.03
	(0.02)			(0.02)	(0.02)	(0.28)			(0.33)	(0.32)	(0.02)			(0.02)	(0.02)
ELA	0.05*			0.07**	0.05*	0.78+			1.20**	0.79+	0.02			0.04	0.02
	(0.02)			(0.02)	(0.02)	(0.40)			(0.41)	(0.39)	(0.02)			(0.02)	(0.02)
N	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340
Adj. R^2	0.208	0.151	0.205	0.175	0.209	0.292	0.220	0.291	0.246	0.294	0.317	0.269	0.315	0.290	0.316
AIC	1274.54	1366.27	1279.31	1328.84	1275.11	8786.04	8913.92	8787.70	8871.63	8784.48	1033.45	1121.38	1037.19	1084.28	1036.30
BIC	1321.35	1402.67	1326.12	1375.64	1332.31	8832.85	8950.32	8834.50	8918.43	8841.68	1080.25	1157.78	1083.99	1131.09	1093.51

Notes: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, cognitive and noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student ever received free- or reduced-price lunch, ever had an individualized education plan, and ever was an English language learner. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all students within each grade in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Online Appendix

Overview

This online appendix includes additional analyses to address three issues from the primary results and sample. We note in the paper that 1) there is evidence of selection in the sample, 2) we can only leverage the self-regulation construct for the noncognitive baseline measure, and 3) we average the noncognitive and cognitive measures rather than use IRT theta scores.

1) To assess sample selection in the main models, we expand the analysis sample to the entire district's 8^{th}-grade students. While we do not have the cognitive and noncognitive data for all 8^{th}-grade students in the district, we do have absences, suspensions, GPA, and test scores. We replicate analyses for college outcomes in **Table A1**. The magnitudes of the effects remain stable, though now, with the larger sample, the standard errors become smaller, and some of the effects that were not statistically significant (e.g., 8th-grade math test scores) are significant in these models. This suggests that the primary results are unrelated to the smaller, higher-performing sample.

2) We explore another more recent sample (Sample 2) with additional noncognitive/socioemotional measures to assess any notable differences when including other measures such as social awareness or self-efficacy. A disadvantage of this sample is that the analysis only follows students through college enrollment. In addition, the sample is smaller and comes only from public charter schools. However, the additional measures provide a clear benefit.

To create this dataset, we combine rich student-level survey data with state administrative records for two cohorts of 8th-grade students attending several public charter middle schools from the same Northeast district as the primary sample during the 2015 and 2016 school years. Within these schools, we sampled all students from whom we received parental consent to participate and who attended school on the data collection day. Students completed surveys assessing their socioemotional development in their regular classrooms. We then merged these data with student-level administrative data—GPA, attendance, suspensions, math and English language arts (ELA) test scores from the state standardized test, and typical demographic information. Scaled scores were standardized by grade, subject, and year by all students in the state to have mean zero and variance one. The administrative data is available for all students, but the student response rates for the survey averaged 73%. We compared the students who responded to the survey and the students who did not respond in **Table A2**. On average, these non-responding students performed worse academically and were likelier to be absent or suspended. Notably, most students in the sample are Black (38%) and Hispanic (48%) and receive free or reduced-price lunches (92%) but are high achieving, performing 0.53 and 0.26 standard deviations above the math and ELA state averages.

We derive our socioemotional data from online and paper surveys to measure socioemotional skills in multiple areas. These surveys were administered to students within their classrooms near the end of the academic year. The survey consisted of seven constructs scales: *(1) self-management*, the ability to regulate one's emotions, thoughts and behaviors with others; (2) *growth mindset*, the belief that one's intelligence is malleable and can grow with effort (3) *grit*, a combination of passion and persistence over an extended period; (4) *social awareness*, the ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports (5 & 6) *self-efficacy in math and ELA*, the belief in one's ability to complete a task or a achieve a goal in math or ELA, respectively; and (7) *mindfulness*, the ability to pay attention in a particular way, on purpose, in the present moment and nonjudgmentally (Kabat-Zinn, 1994). To measure student's grit, we use the Short Grit Scale (Farrington, 2013; Duckworth & Quinn, 2009). Students respond to four items (e.g., "I finish whatever I begin") on a five-point scale ranging from "not like me at all" to "very much like me." To measure mindfulness, we use a short form of the Mindful Attention Awareness Scale (MAAS) adapted for children and adolescents (Black et al., 2012). Students responded to 6 items, and prior work has demonstrated strong reliability (Cronbach's alpha = 0.89-0.93; test-retest r = .35-.52). It included questions such as "I rush through activities without being really attentive to them and answered them on a 6-point scale (*1: almost never, to 6: almost always, reverse coded in for this item*). For the other scales, refer to Meyer et al. (2018) for further details. For all the scales, we take the average score for the items. A validation study using our survey measures, except grit and mindfulness, found the measures to have high structural validity and reliability Meyer et al. (2018).

In **Table A3**, we replicate the high school graduation, high school dropout, and college enrollment analysis without cognitive measures and use self-management measures instead of the self-regulation index using Sample 2. The self-management measure used in the analysis also uses a subset of the Impulsivity Scale for Children (Tsukayama et al., 2013) included in the self-regulation measure from Sample 1. While we lose power in this analysis (i.e., some of the indicators lose statistical significance due to this smaller sample), many patterns we observe from Sample 1 hold. Survey measures are less predictive than the administrative data; GPA is still the strongest predictor; and the absences-suspensions index is predictive (though not for college enrollment in Sample 2). It is also worth noting that the adjusted R-squared and fit statistics indicate that these models explain less variation than Sample 1.

In **Table A4**, we replicate the high school graduation and college enrollment analysis and explore the other types of noncognitive/socioemotional measures added to the model. Models (1) and (7) included only self-management; models (2) and (4) replicate the results of columns (10) and (15) in **Table A3**. Columns (3) and (9) only include survey measures; models (4) and (10) add the administrative measures. Finally, because of the high correlation between several measures except for self-management, growth mindset, and mindfulness, we created an index of all the other noncognitive competencies. Including other noncognitive measures in the model does not add much predictive power (e.g., AIC reduces from 860.07 to 854.16 for high school graduation), and the only significant measure is grit, which has a surprising negative value. When we explored the correlation, grit and self-management correlated by 0.56 in our sample, the odd relationship could be because there could be a negative relationship once the self-management element of grit is held constant. However, broadly in line with prior literature (Farrington et al., 2012), academic persistence is the most important measure for graduation and college enrollment. Thus, self-management/self-regulation remains a stronger predictor of student outcomes than other noncognitive measures for these additional commonly used measures.

3) Finally, the analysis of Sample 1 only included students with complete survey responses, but most students responded to the survey. We used an item response theory (IRT) graded response model (GRM) on the surveys as a specification check. The correlation between the GRM and PCA indices is 0.97. **Table A5** shows the results of Table 5 in the main paper using the GRM measures for the self-regulation index. These results do not differ between the two approaches, so we report the PCA-based index results as the primary results to be comparable to prior research in this area.

References

Black, D. S., Sussman, S., Johnson, C. A., & Milam, J. (2012). Trait mindfulness helps shield decision-making from translating into health-risk behavior. *Journal of Adolescent Health*, 51, 588–592. https://doi.org/10.1016/j.jadohealth.2012.03.011

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.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Meyer, R. H., Wang, C., & Rice, A. B. (2018). Measuring Students' Social-Emotional Learning among California's CORE Districts: An IRT Modeling Approach. Working Paper. *Policy Analysis for California Education, PACE*.

Kabat-Zinn, J. (1994) *Mindfulness meditation for everyday life*. London, UK: Piatkus.

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. Developmental Science, 16, 879–893.

Tables & Figures

Table A1. Predicting College Outcomes Using 8th-grade Administrative Measures for the Full District (Sample 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	College Enrollment			College Quarters			Bachelor's Degree		
Abs-Susp Index	-0.06***		-0.06***	-0.81***		-0.83***	-0.01*		-0.01*
	(0.01)		(0.01)	(0.11)		(0.11)	(0.01)		(0.01)
GPA	0.11***		0.10***	2.27***		2.06***	0.10***		0.09***
	(0.02)		(0.02)	(0.19)		(0.17)	(0.01)		(0.01)
Math		0.07***	-0.00		1.42***	-0.01		0.09***	0.04***
		(0.02)	(0.01)		(0.22)	(0.18)		(0.01)	(0.01)
ELA		0.07***	0.05***		1.27***	0.79***		0.03*	0.01
		(0.01)	(0.01)		(0.20)	(0.19)		(0.01)	(0.01)
N	3723	3723	3723	3723	3723	3723	3723	3723	3723
Adjusted R-squared	0.259	0.203	0.264	0.352	0.281	0.356	0.301	0.267	0.304
AIC	3798.59	4069.56	3777.15	24185.23	24571.07	24163.76	2503.15	2679.86	2487.85
BIC	3842.14	4113.11	3833.15	24228.78	24614.63	24219.76	2546.71	2723.42	2543.85

Notes: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student received free- or reduced-price lunch, had an individualized education plan, or was an English language learner in 8th-grade. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. All models also include school fixed effects. The test scores are standardized across all 8th-grade students in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table A2: Mean Demographic Characteristics and Measures for all 8th-grade Students Among Sampled Schools and Sampled Students (Sample 2)

	(1)	(2)
	All 8th-grade Students in Schools	Students with Complete Data
Male	0.479	0.468
African American	0.393	0.379
Asian/Pacific Islander	0.017	0.018
Hispanic	0.462	0.479
White, Non-Hispanic	0.109	0.105
Free/reduced-price lunch	0.918	0.922
Individualized education plan	0.231	0.197
English language learner	0.265	0.243
Absences (number)	7.803	6.469
Suspension (ever)	0.171	0.156
GPA (standardized)	0.000	0.063
Math scores (standardized)	0.419	0.530
ELA scores (standardized)	0.196	0.263
N	1419	1041

Notes: All samples are restricted to students with complete demographic and academic information. All students could participate in the noncognitive surveys; students with complete data have valid data on the noncognitive subtests. 8th-grade GPA is standardized within the school and across all students in the sample. The test scores are standardized across all 8th-grade students in the state. 8th-grade suspension is a binary for whether the student was ever suspended that year. Absences = the number of absences. GPA = grade point average. ELA = English language arts.

Table A3. Predicting High School Dropout, High School Graduation, & College Enrollment Using 8th-grade Self-Management Survey and Administrative Measures (Sample 2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	High School Dropout					High School Graduation					College Enrollment				
Self-Manag.		-0.00	0.01		0.01		0.03+	0.01		0.01		0.02	-0.00		-0.00
		(0.00)	(0.00)		(0.00)		(0.01)	(0.01)		(0.01)		(0.02)	(0.02)		(0.02)
Abs-Susp Index	0.02+		0.02+		0.02+	-0.03+		-0.03+		-0.03+	-0.02		-0.02		-0.02
	(0.01)		(0.01)		(0.01)	(0.02)		(0.02)		(0.02)	(0.02)		(0.02)		(0.02)
GPA	-0.03***		-0.02*		-0.03**	0.05+		0.03		0.05+	0.09***		0.09**		0.09**
	(0.01)		(0.01)		(0.01)	(0.02)		(0.02)		(0.02)	(0.02)		(0.02)		(0.02)
Math	0.00			-0.02*	0.00	-0.00			0.04	-0.00	0.04			0.10*	0.04
	(0.01)			(0.01)	(0.01)	(0.03)			(0.03)	(0.03)	(0.04)			(0.04)	(0.04)
ELA	0.02*			0.01	0.02*	-0.04*			-0.02	-0.04*	-0.04+			-0.01	-0.04+
	(0.01)			(0.01)	(0.01)	(0.02)			(0.02)	(0.02)	(0.02)			(0.02)	(0.02)
N	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041
Adj. R^2	0.039	0.012	0.035	0.016	0.040	0.081	0.070	0.079	0.068	0.082	0.082	0.053	0.080	0.068	0.081
AIC	-716.29	-688.12	-710.63	-691.24	-717.53	848.04	860.07	851.54	862.78	846.91	1299.07	1330.28	1301.57	1314.04	1299.03
BIC	-661.87	-643.59	-656.20	-641.76	-663.10	902.47	904.60	905.97	912.26	901.34	1353.49	1374.81	1355.99	1363.51	1353.46

Notes: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student received free- or reduced-price lunch, had an individualized education plan, or was an English language learner in 8th-grade. The abs-susp index is oppositely signed to the other measures by construction. A higher value on this index indicates worse outcomes. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all 8th-grade students in the state. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table A4. Predicting High School Graduation and College Enrollment Using 8th-grade Additional Noncognitive Surveys and Administrative Measures (Sample 2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	High School Graduation						College Enrollment					
Self-Manag.	0.03+	0.01	0.04+	0.03	0.04+	0.03	0.02	-0.00	0.03	0.01	0.02	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Growth Mindset			0.01	0.01	0.02	0.01			0.03	0.01	0.03	0.01
			(0.01)	(0.01)	(0.01)	(0.01)			(0.02)	(0.02)	(0.02)	(0.02)
Grit			-0.03+	-0.04*					-0.01	-0.02		
			(0.02)	(0.02)					(0.02)	(0.02)		
Self-Eff-Math			0.01	0.00					0.02	-0.02		
			(0.01)	(0.01)					(0.01)	(0.01)		
Self-Eff-ELA			-0.01	-0.00					0.01	0.03		
			(0.02)	(0.02)					(0.02)	(0.02)		
Social Awareness			0.00	0.00					-0.03	-0.02		
			(0.02)	(0.02)					(0.02)	(0.02)		
Mindfulness			0.02	0.01	0.02	0.01			0.01	0.00	0.01	-0.00
			(0.01)	(0.01)	(0.01)	(0.01)			(0.01)	(0.01)	(0.01)	(0.01)
Abs-Susp Index					-0.03+	-0.03*					-0.01	-0.02
					(0.01)	(0.01)					(0.02)	(0.02)
GPA		-0.03+		-0.03		-0.03		-0.02		-0.02		-0.02
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
Math		0.05+		0.05+		0.05+		0.09**		0.09***		0.09***
		(0.02)		(0.03)		(0.02)		(0.02)		(0.02)		(0.02)
ELA		-0.00		-0.01		-0.00		0.04		0.05		0.03
		(0.03)		(0.02)		(0.03)		(0.04)		(0.04)		(0.04)
SEL Index - No GM, SM, Mindful		-0.04*		-0.04*		-0.04*		-0.04+		-0.06*		-0.04+
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
N	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041	1041
Adj. R^2	0.070	0.082	0.073	0.083	0.072	0.083	0.053	0.081	0.055	0.081	0.055	0.079
AIC	860.07	846.91	854.16	838.74	858.19	842.13	1330.28	1299.03	1325.55	1292.84	1329.20	1297.97
BIC	904.60	901.34	908.59	893.16	912.62	896.56	1374.81	1353.46	1379.98	1347.26	1383.62	1352.40

Notes: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student received free- or reduced-price lunch, had an individualized education plan, or was an English language learner in 8th-grade. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all 8th-grade students in the state. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1

Table A5. Predicting College Outcomes Using 8th-grade GRM-based Self-Regulation Index, Cognitive Index, and Administrative Measures (Sample 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	College Enrollment					College Quarters					Bachelor's Degree in Four Years				
Self-Reg Index		0.03**	0.01	0.03*	0.01		0.79***	0.23*	0.66***	0.24*		0.04***	0.01	0.03***	0.01
		(0.01)	(0.01)	(0.01)	(0.01)		(0.15)	(0.10)	(0.14)	(0.10)		(0.01)	(0.01)	(0.01)	(0.01)
Cog Index		0.01	-0.01	-0.02	-0.02		0.19	-0.27	-0.26	-0.32+		0.03*	0.00	-0.00	-0.00
		(0.01)	(0.01)	(0.01)	(0.01)		(0.18)	(0.16)	(0.21)	(0.19)		(0.01)	(0.01)	(0.01)	(0.01)
Abs-Susp Index	-0.05***		-0.05***		-0.05***	-1.11***		-1.09***		-1.12***	-0.03+		-0.03+		-0.03+
	(0.01)		(0.01)		(0.01)	(0.17)		(0.17)		(0.17)	(0.01)		(0.01)		(0.01)
GPA	0.08**		0.10***		0.08***	1.86***		1.93***		1.69***	0.09***		0.10***		0.09***
	(0.02)		(0.02)		(0.02)	(0.30)		(0.34)		(0.29)	(0.02)		(0.01)		(0.02)
Math	-0.00			0.06*	0.01	-0.22			1.09**	0.02	0.03			0.08***	0.03
	(0.02)			(0.02)	(0.02)	(0.28)			(0.33)	(0.32)	(0.02)			(0.02)	(0.02)
ELA	0.05*			0.06**	0.05*	0.78+			1.17**	0.78+	0.02			0.03	0.02
	(0.02)			(0.02)	(0.02)	(0.40)			(0.41)	(0.39)	(0.02)			(0.02)	(0.02)
N	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340	1340
Adj. R^2	0.208	0.153	0.205	0.177	0.209	0.292	0.224	0.292	0.248	0.295	0.317	0.273	0.315	0.293	0.317
AIC	1274.54	1361.99	1278.96	1325.99	1274.69	8786.04	8907.07	8786.58	8866.65	8783.35	1033.45	1114.81	1036.10	1079.43	1035.06
BIC	1321.35	1398.39	1325.77	1372.79	1331.90	8832.85	8943.47	8833.39	8913.45	8840.55	1080.25	1151.22	1082.91	1126.23	1092.27

Notes: Standard errors are reported in parentheses and clustered at the 8th-grade middle school attended. Samples are restricted to students with complete demographic, cognitive, and noncognitive surveys, 8th-grade behavioral and academic outcome data, and a non-missing outcome. All regressions control for student gender and race/ethnicity and for whether a student received free- or reduced-price lunch, had an individualized education plan, and was an English language learner. All models also include school fixed effects. All survey measures are standardized within the sample of students who completed the survey. The test scores are standardized across all students within each grade in the district. *** p<0.001, ** p<0.01, * p<0.05, + p< 0.1