EdWorkingPaper No. 25-1250

Measuring Academic Behaviors at Scale: Developing Longitudinal Measures of "Noncognitive" Skills Using Administrative Data and Examining Predictive Validity

Tiffany Wu
University of Michigan

Christina Weiland
University of Michigan

Matthew A. Diemer
University of Michigan

Rebecca Unterman
MDRC

Anna Shapiro
RAND Corporation

Thomas Staines
University of Michigan

Academic behaviors, as reflected in administrative indicators such as attendance and suspensions, are widely used in education research. However, limited guidance exists on how to measure them longitudinally and at scale. Using longitudinal data from Boston Public Schools students followed from kindergarten through high school (N = 12,232), we compare grade-specific point-in-time indicators with longitudinal composites constructed using factor analysis and principal components analysis (PCA). Factor analyses suggest that academic behaviors are multidimensional, reflecting distinct but correlated patterns of attendance and disciplinary behaviors across developmental periods. Despite this structure, PCA-based and factor-analytic composites demonstrated broadly similar predictive validity for eighth grade standardized test scores, on-time high school graduation, and immediate college enrollment. Eighth grade unexcused absence rate, a driver of both composites, performed comparably to longitudinal composites and also predicted later educational attainment above and beyond prior achievement. Findings suggest that, when prediction is the goal, longitudinal composites may yield limited incremental benefits over carefully selected point-in-time attendance measures. We offer practical guidance for researchers and policymakers seeking scalable, interpretable measures of student behavior to inform interventions and policies.

VERSION: August 2026

Suggested citation: Wu, Tiffany, Christina Weiland, Matthew A. Diemer, Rebecca Unterman, Anna Shapiro, and Thomas Staines. (2026). Measuring Academic Behaviors at Scale: Developing Longitudinal Measures of "Noncognitive" Skills Using Administrative Data and Examining Predictive Validity. (EdWorkingPaper: 25-1250). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/7h7f-0j56

Measuring Academic Behaviors at Scale:

Developing Longitudinal Measures of "Noncognitive" Skills Using Administrative Data and Examining Predictive Validity

Tiffany Wu [1], Christina Weiland [1], Matthew A. Diemer[1], Rebecca Unterman [2], Anna Shapiro [3], Thomas Staines [1]

[1]University of Michigan, [2]MDRC, [3]RAND

Author's note: This study was supported by grants from the Institute of Education Sciences, U.S. Department of Education: R305A220036, R305B20001, and R305B200011. We thank Jason Sachs, the Boston Public Schools (especially TeeAra Dias, Yuzhu Xia, and Megan Toney), and the Massachusetts Department of Elementary and Secondary Education (especially Brendan Longe, Matthew Deninger, and Kendra Winner) for their partnership, feedback, and support. ChatGPT (version 5.1) was used to assist with troubleshooting programming issues, improving code efficiency, and proofreading. All substantive arguments, interpretations, and conclusions were developed independently by the authors. Any AI-generated content was critically reviewed and revised to ensure accuracy and consistency with the study's objectives. Correspondence concerning this article should be addressed to Tiffany Wu, Marsal Family School of Education, 610 E. University Ave, Ann Arbor, MI 48104, or via e-mail at wutiffa@umich.edu.

Abstract

Academic behaviors, as reflected in administrative indicators such as attendance and suspensions, are widely used in education research. However, limited guidance exists on how to measure them longitudinally and at scale. Using longitudinal data from Boston Public Schools students followed from kindergarten through high school (N = 12,232), we compare grade-specific point-in-time indicators with longitudinal composites constructed using factor analysis and principal components analysis (PCA). Factor analyses suggest that academic behaviors are multidimensional, reflecting distinct but correlated patterns of attendance and disciplinary behaviors across developmental periods. Despite this structure, PCA-based and factor-analytic composites demonstrated broadly similar predictive validity for eighth grade standardized test scores, on-time high school graduation, and immediate college enrollment. Eighth grade unexcused absence rate, a driver of both composites, performed comparably to longitudinal composites and also predicted later educational attainment above and beyond prior achievement. Findings suggest that, when prediction is the goal, longitudinal composites may yield limited incremental benefits over carefully selected point-in-time attendance measures. We offer practical guidance for researchers and policymakers seeking scalable, interpretable measures of student behavior to inform interventions and policies.

Key words: academic behavior, PCA, factor analysis, psychometrics, noncognitive skills

Measuring Academic Behaviors at Scale:

Developing Longitudinal Measures of "Noncognitive" Skills Using Administrative Data and Examining Predictive Validity

Student success in school requires a range of skills beyond those measured by standardized testing. These non-test score outcomes are often grouped under the broad umbrella of "noncognitive" skills. Among the five common "noncognitive" domains (Farrington et al., 2012), *academic behaviors*—observable behaviors that are "commonly associated with being a 'good student'" (Farrington et al., 2012, p. 1) such as regularly attending class and avoiding disciplinary infractions—have become central to education research and policy. Academic behaviors act as observable signs of student engagement and are strongly associated with academic achievement and longer-term outcomes (Cleveland & Scherer, 2025; Liu et al., 2025).

Academic behaviors can be measured in many ways, but education researchers are increasingly operationalizing these constructs using administrative data routinely collected in schools, such as student attendance and disciplinary records (Backes et al., 2022; Cleveland & Scherer, 2025). The Every Student Succeeds Act's (ESSA) requirement that states include at least one non-test-score indicator of student success in their accountability systems has increased the prominence of administrative measures of academic behavior. Recent longitudinal studies and intervention research, including those supported by the Institute of Education Sciences (IES), have tracked these outcomes over multiple years to evaluate program impacts (e.g., Gray-Lobe et al., 2023; Weiland et al., 2025), and school-based early warning systems and on-track frameworks monitor these same behavioral indicators over time to identify students at risk of adverse educational outcomes (e.g., Allensworth, 2013; Wu et al., 2026). Prior work has shown that administrative indicators of academic behaviors can be more predictive of outcomes such as on-time high school graduation and college enrollment compared to survey-based social and emotional learning (Liu et al., 2025) and self-regulation and cognition measures (Cleveland & Scherer, 2025).

Despite the growing use of administrative indicators for analyzing students' behavioral trajectories, relatively little guidance exists on how to measure academic behaviors longitudinally and at scale. Researchers conducting longitudinal studies, for example (e.g., Weiland et al., 2025), often track multiple behavioral indicators across multiple years, fitting separate impact models for each one, but this raises concerns about multiple comparisons. Developing composite measures offers one path forward by synthesizing information across indicators and years into a single metric. However, there have been substantial researcher degrees of freedom in how academic behavior indicators are combined, including which administrative indicators and grade levels are included, and different disciplines emphasize different measurement approaches. For example, psychologists often rely on factor analytic methods whereas economists have frequently used principal components analysis (PCA) to create composite indices (Furr, 2021; Heckman et al., 2006).

Moreover, it remains unclear whether longitudinal composites could outperform grade-specific point-in-time indicators (e.g., a student's eighth grade unexcused absence rate) in predicting long-term student outcomes. If point-in-time measures are equally informative, the additional complexity of constructing longitudinal composites yield limited predictive benefit. Conversely, if longitudinal composites better capture persistent patterns, they may offer advantages for applications such as early warning systems. More guidance is needed to evaluate the trade-offs among different approaches and to identify best practices for constructing reliable measures that facilitate comparisons across studies.

In this paper, we contribute to these open questions by using a demographically diverse sample of students followed from kindergarten to 12th grade in Boston Public Schools (N = 12,232) to compare three approaches to measuring academic behaviors longitudinally using administrative data: grade-specific point-in-time indicators, factor analysis, and principal components analysis (PCA). We discuss the psychometric properties of the latter two approaches and evaluate the predictive validity of all three using students' eighth grade standardized test scores, on-time high school graduation, and immediate college enrollment. The sample spans cohorts whose educational experiences both preceded and were affected by the COVID-19 pandemic, with outcome data available through the 2023-2024 school year. This study bridges two key needs in the field: supporting rigorous, comparable evaluations of academic behaviors and enhancing the utility of predictive indicators used by schools and districts.

Approaches to Measuring Academic Behaviors Using Administrative Data

Point-in-time Indicators. Point-in-time indicators, such as a student's third grade unexcused absence rate, are a common approach to measuring academic behaviors in education research. Because they are derived directly from routinely collected administrative records, they are straightforward to construct, easy to interpret, and closely aligned with how schools monitor student progress (Liu et al., 2025; Weiland et al., 2025). However, they may be sensitive to temporary fluctuations and fail to reflect broader patterns of behavior across years of schooling. Reliance on multiple point-in-time indicators can also increase concerns about multiple comparisons, yet averaging across years raises additional questions about which grade ranges to combine and how.

Principal Components Analysis (PCA). One approach is to combine multiple academic behavior indicators into a composite measure using PCA, a statistical technique used to reduce the dimensionality of a dataset by identifying a smaller set of uncorrelated components that capture the most variance across the original variables (Furr, 2021). PCA can reduce the number of outcomes examined and mitigate multiple-comparison concerns. Its emphasis on maximizing explained variance may also make it particularly useful when the primary goal is prediction. However, PCA does not distinguish common variance from measurement error or unique variance, and the resulting components may not correspond to theoretically meaningful dimensions of academic behavior (Floyd & Widaman, 1995; Preacher & MacCallum, 2003).

Factor Analysis. Factor analysis has not been widely applied to longitudinal administrative data in this context but holds potential to address some of the shortcomings of PCA. To our knowledge, only one prior study constructed a "nontest index...using a factor model" (Backes et al., 2022, p. 21). Unlike PCA, factor analysis explicitly models the latent constructs underlying observed variables, making it well-suited for creating a theoretically grounded student behavior measure (Furr, 2021). Because it parses out measurement error in administrative data, it could provide more precise estimates of academic behavior. This approach also allows for the inclusion of correlated factors, which aligns with the theoretical understanding that student behaviors like attendance and discipline are interrelated (PCA produces orthogonal components by construction). However, factor analytic approaches require a series of subjective modeling decisions, including the number of factors to retain and the rotation method. It remains unclear whether theoretically grounded measures provide advantages over point-in-time indicators or PCA-based composites for predicting student outcomes.

Current Study

To examine trade-offs between these three focal approaches for measuring academic behaviors at scale and longitudinally, we examine the following research questions:

1. What latent structure of academic behavior does factor analysis uncover, using longitudinal administrative data from kindergarten to eighth grade?
2. What structure of academic behavior does PCA uncover, using longitudinal administrative data from kindergarten to eighth grade? How does the structure compare to a factor analytic approach?
3. How do factor analytic measures, PCA-based measures, and grade-specific point-in-time indicators of academic behavior differ in their predictive validity for eighth grade end-of-year math and ELA standardized test scores?
4. How do factor analytic measures, PCA-based measures, and standalone grade-specific indicators of academic behavior differ in their predictive validity for on-time high school graduation and immediate college enrollment, controlling for eighth grade standardized test scores?

We constructed factor analytic and PCA-based composite measures using kindergarten through eighth grade data (and not covering grades 9 through 12) for two reasons. First, eighth grade marks the end of middle school and precedes the transition to high school, making it a substantively important endpoint for measuring academic behaviors (Eccles et al., 1993). Second, the COVID-19 pandemic disrupted traditional schooling beginning in spring 2020, when students in our cohorts were in grades 8-11, followed by a year of virtual and later hybrid instruction when students were in grades 9-12 (Appendix A Table 1). During the virtual learning year, the Boston Public Schools district deemed attendance records, a key measure of academic behaviors, from the virtual learning year unreliable for research purposes. Accordingly, K-8 provided the most consistent longitudinal follow-up period for our measurement aims.

Method

Sample

Our sample was the population of students who applied to the Boston Public Schools (BPS) Pre-K program for four-year-olds between the 2007-2008 and 2010-2011 school years. This sample is the focus of an ongoing longitudinal impact study of the Boston Pre-K program's impacts (Weiland et al., 2025). We tracked students from their Pre-K application through high school for each focal cohort (up to school year 2023-2024; Appendix A Table 1). However, we did not use data from the Pre-K year as not all applicants in the sample enrolled in Pre-K and thus we lacked data on Pre-K academic behaviors. Our full sample consisted of 12,232 students. On average, this sample was 52% male and ethnically and racially diverse (8% Asian, 28% Black, 44% Hispanic, 3% multiracial or other, and 17% White). More than half the students in the sample qualified for free or reduced-price lunch (68%), 93% of the students were born in the United States, and 39% of the students had a first language other than English.

Measures

We examined a set of administrative indicators for creating longitudinal academic behavior composites: unexcused absence rate, chronic absenteeism, number of suspensions, whether a student was ever suspended in a given grade, and grade retention each year from kindergarten to eighth grade. To do so, we merged student-level records from the Student Information Management System (SIMS) and the School Safety and Discipline Report datasets provided by the Massachusetts Department of Elementary and Secondary Education. To examine predictive validity, we used graduation records from SIMS and further linked these records to test score data from the Massachusetts Comprehensive Assessment System (MCAS) and postsecondary enrollment data from the National Student Clearinghouse. Descriptive statistics on these indicators are presented in Table 1.

Unexcused Absence Rate and Chronic Absenteeism. We calculated the unexcused absence rate variables as the number of unexcused absence days divided by the total number of days a student was enrolled in their primary school for a given year. We used a rate rather than the raw count of unexcused absences to account for variations in enrollment days both between public and charter schools (Wu et al., 2026). We focused on unexcused absences rather than total absences since excused absences can reflect reasons for missing school, such as illness, that do not directly reflect student behavior. Consistent with Jackson (2018), we took the log of the unexcused absence rate (plus 1) during our factor analyses to account for its right-skewed distribution. We also constructed binary indicators for chronic absenteeism, with a value of one indicating that a student had a total absence rate of 10% or more in a given school year, following the most common definition of this construct (Wu et al., 2026).

Suspensions. We created an indicator for the total number of suspensions a student received in a given school year, including both in-school and out-of-school suspensions. We combined these categories because the Succeed Boston restorative justice program is administratively classified as an out-of-school suspension despite operating similarly to an in-school suspension program, making distinctions between suspension types less meaningful in this context (Succeed Boston, 2025; see Appendix B). We took the log of this variable (plus 1) during our factor analyses to account for its right-skewed distribution. We also created binary K-8 measures of whether a student was ever suspended during a given school year (regardless of how many times they were suspended), with a value of one indicating that they had.

Grade Retention. We created binary indicators of grade retention, with a value of one indicating that a student was retained for a given grade. We chose to include grade retention as an academic behavior indicator because it represents a school's behavioral response to a student's overall academic and engagement trajectory. In this sense, retention can capture the institutional consequences of chronic disengagement.

Eighth Grade Standardized Test Scores. For our third and fourth research questions, we used students' 8th grade math and English language arts (ELA) standardized test scores from the MCAS statewide test. We standardized each student's theta (i.e., IRT-derived) score on the mean and standard deviation of all eighth graders within Boston Public Schools taking the exam the same year. This standardization followed guidance from the Massachusetts Department of Elementary & Secondary Education (2016) to facilitate longitudinal score comparisons.

On-time High School Graduation. We created a binary indicator for on-time high school graduation based on students' expected educational progression from Pre-K, defined as graduating from high school by the year they would be expected to do so in the absence of grade retention.

Immediate College Enrollment. We created a binary indicator for immediate college enrollment, defined as enrolling in a postsecondary institution within one year of a student's expected high school graduation year.

Covariates. We constructed a set of student-level covariates for the fourth research question. We captured students' race/ethnicity using a set of binary variables for whether student identified as Asian, Black, Hispanic, multiracial/other, or White. Similarly, we created a set of binary variables to identify whether a student's first language was English, whether a student's country of origin was the United States, and whether the student was eligible for free- or reduced-priced lunch. We also calculated students' age as of September 1 in the year they were expected to enter eighth grade, based on their pre-kindergarten application year.

Data Analytic Approach

Following the scale validation literature, we first conducted an exploratory factor analysis (EFA) followed by a confirmatory factor analysis (CFA) (Worthington & Whittaker, 2006). We divided the larger sample (N = 12,232) into two randomly selected and independent (i.e., without replacement) datasets of equal size (N = 6,116 each). The first dataset was used to establish the factor structure through EFA while the second dataset was used to validate this structure through CFA. The primary EFA approach used maximum likelihood (ML) extraction method run under full information maximum likelihood (FIML) conditions to address missing data (Bates et al., 2019). However, results were robust across alternative extraction methods and estimators. For more information on the robustness analyses, along with detailed information on factor retention decisions, factor loadings, model fit indices, and reliability statistics, see Appendix C. After the EFA step, we dropped indicators (i.e., grade retention, chronic absenteeism, and binary variable on whether a student was ever suspended) that had low loadings. Our primary CFA approach used maximum likelihood estimation with robust standard errors (MLR), with estimated correlations between the factors and missing data handled using FIML. Results were robust across alternative estimation and missing data approaches (Appendix D).

We conducted PCA using the full set of administrative academic behavior indicators included in the EFA (Table 1). All indicators were standardized prior to analysis, and, consistent with prior education research (e.g., Heckman et al., 2006; Jackson, 2018), we retained the first principal component as the composite index. Detailed PCA loadings, scree plots, proportion of variance explained, and robustness checks are reported in Appendix E.

To answer the third and fourth research questions on predictive validity, we estimated a series of regression models using the larger sample (N = 12,232). Specifically, we modeled student outcomes as a function of each academic behavior measure, a set of student demographic covariates, and a school-level random intercept:

$$Y_{ij} = \beta_0 + \beta_1 AcademicBehavior_{ij} + \boldsymbol{\gamma}'\mathbf{X}_{ij} + u_j + \varepsilon_{ij}, \quad u_j \sim N(0, \tau^2)$$

The primary parameter of interest is β_1, which represents the relationship between the academic behavior measure and the outcome. For the third research question, Y_{ij} represents either the eighth grade standardized math or ELA score for student i in school j, $AcademicBehavior_{ij}$ represents the focal academic behavior measure, $\mathbf{X}_{ij}$ is a vector of student demographic covariates, u_j is a school-specific random intercept for the student's eighth grade school, and e_{ij} is the student-level residual error term.

For the fourth research question, Y_{ij} represents either a binary indicator for on-time high school graduation or immediate college enrollment for student i in school j. $AcademicBehavior_{ij}$ and all other terms are defined as above. For these models, we additionally included students' eighth grade math and ELA standardized test scores in addition to the student demographic covariates of $\mathbf{X}_{ij}$ to assess whether academic behavior measures predicted long-term educational outcomes above and beyond prior academic achievement.

Results

Factor Analysis

EFA identified a four-factor structure that provided the best balance of statistical fit and substantive interpretability. As seen in Table 2, the retained factors were divided by behavioral domain and grade band: (a) primary grades unexcused absences ("K-2 Unexcused Absence"); (b) elementary and middle grades unexcused absences ("3-8 Unexcused Absence"); (c) early grades suspensions ("2-4 Suspensions"); and (d) middle grades suspensions ("5-8 Suspensions"). During this stage, variables in Table 1 with weak loadings (i.e., grade retention, chronic absenteeism, ever suspended) were removed. The final structure suggests that academic behaviors are best characterized by correlated dimensions of attendance and discipline that vary across grades rather than a single underlying construct. This may reflect developmental changes, different school structures or reporting practices across grades, and/or the low prevalence of some indicators in early grades.

We cross-validated the four-factor structure using CFA. All factor loadings were statistically significant and exceeded 0.40 in magnitude (Table 2). The CFA reproduced the four-factor relationships observed in the EFA and demonstrated excellent model fit (all robust CFI and TLI values > 0.95, RMSEA < 0.05, SRMR ≈ 0.03; Bentler, 1990; Hu & Bentler, 1999), providing strong evidence that longitudinal academic behavior is best characterized by four correlated factors capturing distinct developmental patterns of unexcused absences and suspensions from kindergarten through eighth grade.

To test predictive validity (research questions 3 and 4), we created 10 factor analytic composite measures based on the results of this factor analysis. One set of constructs utilized the factor loadings derived from the CFA as weights for each of the four factors. This set of measures has the terms “FA,” the factor name (e.g., “3-8 Unexcused Absence”), and “Weighted Measure” when referenced in Figure 1. To enhance replicability of our measures, we also created a simpler version of each of the weighted measures by calculating the unweighted mean of the variables in each factor. This set of measures has the terms “FA”, the factor name (e.g., “3-8 Unexcused Absence”), and “Unweighted Measure” in them in Figure 1. Additionally, we constructed both weighted and unweighted overall averages across all the variables in the four factors. These measures have the terms “FA” and “All FA Variables” in Figure 1 and were made to be more comparable to the PCA measure. A higher value (i.e., greater unexcused absences) indicates worse outcomes.

Principal Components Analysis

We constructed five PCA-based composite measures. First, we created a PCA index using the first principal component of all variables for all grades included in our initial EFA models, as listed in Table 1. Additionally, we generated four grade-range PCA indices to account for potential temporal trends: one just using variables from grades K-2, one using variables from elementary and middle grades 3-8, one using variables from grades K-5, and one using measures from grades 6-8. The first two grade groupings were informed by the grade splits identified in the factor analysis for unexcused absences. The latter two groupings reflect conceptual distinctions commonly made between early and later schooling stages.

To stay consistent with the prior work (e.g., Cleveland & Scherer, 2025; Heckman et al., 2006; Jackson, 2018), we kept the first principal component for each composite measure. However, as seen in Appendix E, the PCA results highlighted some important considerations regarding the appropriateness of using the first principal component to summarize academic behavior. The scree plot of the PCA with all variables (Appendix E Figure 1) indicated multiple dimensions with an eigenvalue greater than one and a steep drop in eigenvalues after the first three components, suggesting that three components should be extracted. For the same PCA, the first principal component explained 16.3% of the variance, while the first three components jointly explained 30.4% of the variance. This suggests that the data may not be adequately summarized by one principal component alone. As a robustness check, we re-estimated our predictive validity models for the third research question using the second principal component of each PCA specification, but these measures showed substantially weaker associations with eighth grade test scores than their first component counterparts (Appendix E, Table 2), supporting our decision to retain only the first principal component as our primary PCA-based measure. This choice facilitates direct comparisons with the existing literature while also highlighting a limitation of the PCA approach: a single principal component may provide only a partial summary of the multidimensional patterns of student academic behavior based on the indicators we used. A higher value indicates worse outcomes.

Predictive Validity for Eighth Grade Standardized Tests

Figure 1 presents the standardized associations for each academic behavior measure, providing a comparison of how these measures relate to test score outcomes. Exact standardized estimates for Figure 1 are in Appendix F Tables 1-2, and robustness checks using school fixed effects are in Appendix F Tables 3-4, which showed similar patterns.

The multilevel regressions revealed that several academic behavior measures statistically significantly predicted eighth grade end-of-year ELA standardized test scores, with overall stronger negative standardized associations observed for grades closer to 8th grade (Appendix F Table 1). The PCA measure for grades 6-8 had the strongest standardized association (red dot in row 1 in Figure 1; $\beta = -0.194$, $p < 0.001$), followed closely by the PCA measure for grades 3-8 (red dot in row 2 in Figure 1; $\beta = -0.184$, $p < 0.001$), the unexcused absence rate for eighth grade (red dot in row 4 in Figure 1; $\beta = -0.178$, $p < 0.001$), the PCA of all the variables (red dot in Row 3 in Figure 1; $\beta = -0.176$, $p < 0.001$), and the factor analysis weighted measure using all the factors (red dot in Row 5 in Figure 1; $\beta = -0.170$, $p < 0.001$). For these top five strongest associations, a one standard deviation increase in the student behavior administrative measure was associated with approximately a 0.17-0.19 standard deviation decrease in ELA standardized test score, holding all else constant. Among the earlier grades, standalone indicators related to unexcused absences and suspensions exhibited smaller but still significant associations, with diminishing effects as the grades became more distant from eighth grade.

Similarly, multilevel regressions demonstrated that several student behavior measures statistically significantly predicted eighth grade end-of-year math standardized test scores (Appendix F Table 2). The PCA measure for grades 3-8 had the strongest standardized association with eighth grade end-of-year MCAS math scores (blue square in Row 1 in Figure 1; $\beta = -0.215$, $p < 0.001$). For these top five strongest associations, a one standard deviation increase in the student behavior administrative measure was associated with approximately a 0.20-0.21 standard deviation decrease in math standardized test score, holding all else constant. Among the earlier grades, variables related to unexcused absences and suspensions exhibited smaller but still significant associations, with diminishing effects as the grades became more distant from eighth grade.

Notably, across both the ELA and math regressions, the differences in predictive strength between the strongest PCA and the strongest factor analytic measure were modest, around a 0.01-0.02 difference in the standardized estimate. As seen in Figure 1, many of the confidence intervals for each standardized estimate overlap. The eighth grade unexcused absence rate point-in-time indicator also had similarly strong magnitudes.

Predictive Validity for High School Graduation and Immediate College Enrollment

Tables 3 and 4 present model results predicting on-time high school graduation and immediate college enrollment, respectively, using each of the top five academic behavior measures with the greatest overall predictive validity for eighth grade standardized test scores ("PCA: 6-8", "PCA: 3-8", "PCA: All Variables", "FA: 3-8 Unexcused Absence Unweighted Composite", "Unexcused Absence Rate Grade 8"). For each academic behavior, column (a) reports models including only the academic behavior measure of interest, while column (b) additionally controls for eighth grade math and ELA standardized test scores. Because statewide assessments were canceled during the COVID-19 pandemic, eighth grade test scores were unavailable for our final cohort, resulting in a smaller analytic sample for the models in column (b). As a robustness check, we re-estimated these models using seventh grade math and ELA test scores, which were available for all cohorts; results were substantively unchanged (Appendix G Tables 1-2).

Across all specifications, the top five academic behavior measures were statistically significantly associated with on-time high school graduation. In models excluding standardized test controls, a one standard deviation increase in the longitudinal PCA and factor-analytic composite measures was associated with a 14.1 to 14.3 percentage point (*pp*) decrease in the probability of graduating on time, holding all else constant ($p < 0.001$). The point-in-time eighth grade unexcused absence rate was modestly less predictive but remained strongly associated with graduation, corresponding to an 11.0 *pp* decrease in the probability of on-time graduation ($p < 0.001$). After controlling for eighth grade math and ELA achievement, the magnitude of the academic behavior coefficients decreased modestly (11.5 to 12.4 *pp*) but remained statistically significant.

The top five academic behavior measures were also statistically significantly associated with immediate college enrollment, with patterns similar to those observed for high school graduation. In models excluding standardized test controls, a one standard deviation increase in the longitudinal PCA and factor-analytic measures was associated with a 12.1 to 12.8 *pp* decrease in the probability of immediate college enrollment ($p < 0.001$), whereas the standalone eighth-grade absence rate was somewhat less predictive but remained statistically significant, corresponding to a 9.6 *pp* decrease ($p < 0.001$). After controlling for eighth-grade math and ELA achievement, the magnitude of the academic behavior coefficients decreased modestly (10.1 to 10.8 *pp*) but remained statistically significant.

These results suggest that both longitudinal behavior composites and timely attendance measures provide predictive information beyond standardized test scores. Moreover, differences in predictive validity between the five academic behavior measures were modest. This finding aligns with a growing body of research demonstrating that administrative indicators of academic behavior capture dimensions of student engagement and educational attainment that are not fully reflected in standardized test performance (Cleveland & Scherer, 2025; Jackson, 2018; Liu et al., 2025).

Discussion & Conclusion

Our study compares three different approaches for measuring academic behavior longitudinally and at scale using administrative data. Three findings stand out. First, PCA and factor-analytic composite measures exhibited similar predictive validity across outcomes, and both captured predictive information above and beyond standardized test scores. While some PCA-based measures showed marginally stronger associations with eighth grade standardized test scores, differences between the two approaches were small and largely disappeared when looking at the top five measures in predicting on-time high school graduation and immediate college enrollment. Researchers may favor factor analysis when prioritizing construct validity, theoretical alignment, and interpretability, whereas PCA may be appealing when the primary goal is data reduction and the construction of parsimonious composite measures. Given the modest differences in predictive validity observed across approaches, the choice between PCA and factor analysis ultimately depends more on the research objective than on meaningful differences in predictive performance.

Second, notably, eighth grade unexcused absence rate performed comparably to the longitudinal composites for eighth grade achievement and was also statistically significantly associated with graduation and college enrollment, capturing predictive information beyond prior achievement. This underscores the practical value of recent attendance information, potentially because it captures contemporaneous student engagement and circumstantial information. While PCA and factor analytic measures may provide a more holistic perspective of behavior and capture trends across a student's academic trajectory, researchers should not overlook the potential of point-in-time attendance indicators for practical applications that require timely, parsimonious, and interpretable inputs. This finding is consistent with prior evidence on the predictive power of timely absenteeism data (e.g., Allensworth, 2013; Liu et al., 2025; Wu & Weiland, 2026) and suggests that the predictive strength of behavioral indicators depends on temporal proximity to the outcome, highlighting the importance of aligning measurement windows with the phenomena they are intended to predict.

Third, academic behaviors were better represented by multiple correlated dimensions that differed by behavioral domain and grade range than by a single general construct, suggesting distinct phases of student behavior across time. This finding mirrors broader developmental transitions, though it is possible that changes in reporting practices, institutional differences, or differences in disciplinary enforcement across early elementary to middle school contribute to these patterns. While prior research has found that overall absenteeism patterns tend to be stable in the early grades from Pre-K to third grade (Dubay & Holla, 2016; Wei, 2024), our factor analysis suggests that third grade may mark a developmental or institutional shift in how absenteeism manifests or is recorded. The stronger predictive performance of later-grade attendance indicators and composites that incorporate them suggests that the value of longitudinal measures may depend not only on the amount of information they aggregate but also on whether they preserve information from developmentally proximate periods that are most relevant to the outcome being predicted.

While this study provides valuable insights into the psychometric properties of creating a student behavior measure using administrative data, several limitations should be noted. First, although our predictive validity models included a rich set of demographic covariates, unobserved factors may influence both academic behaviors and later educational outcomes. Consequently, our models run the risk of omitted variable bias and should not be interpreted as causal. Additionally, our models may overstate the contribution of academic behavior measures relative to studies that control for a broader set of academic predictors; for example, Cleveland and Scherer (2025) found smaller associations between their absences-suspension index and later outcomes after including GPA. Our reliance on administrative records also has inherent limitations relative to richer methods like surveys or teacher rating scales, and these measures should be contextualized within the broader structural conditions that shape them, since they can inadvertently encode and perpetuate systemic inequities if interpreted uncritically (Lenhoff & Pogodzinski, 2018).

As education researchers and practitioners turn to administrative data to inform decision-making, it is imperative to recognize both the strengths and the limitations of different academic behavior measures. Moving forward, continuing to evaluate the validity of behavioral measures in different contexts and uses will be essential in developing measures that better support students across their educational journeys.

References

Allensworth, E. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 68-83.

Backes, B., Cowan, J., & Goldhaber, D., & Theobald, R. (2022). *Teachers and school climate: Effects on student outcomes and academic disparities* (CALDER Working Paper No. 274-1022). The National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Bates, T. C., Maes, H., & Neale, M. C. (2019). umx: Twin and path-based structural equation modeling in R. *Twin Research and Human Genetics*, *22*(1), 27-41.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge. https://doi.org/10.4324/9780203807644

Cleveland, C., & Scherer, E. (2025). Are students on track? Comparing the predictive validity of administrative and survey measures of cognitive and noncognitive skills for long-term outcomes. *Educational Researcher*, *54*(4), 213-225.

Dubay L., & Holla N. (2016). *Does attendance in early education predict attendance in elementary school? An analysis of DCPS's early education program.* Urban Institute.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

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*. University of Chicago Consortium on School Research.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Furr, R. M. (2021). *Psychometrics: An introduction*. SAGE publications.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*(1), 363-411.

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.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy, 126*(5), 2072-2107.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kyriazos, T. A., & Poga-Kyriazou, M. (2023). Applied psychometrics: Estimator considerations in commonly encountered conditions in CFA, SEM, and EFA practice. *Psychology, 14*(5), 799-828. https://doi.org/10.4236/psych.2023.145043

Lenhoff, S. W., & Pogodzinski, B. (2018). School organizational effectiveness and chronic absenteeism: Implications for accountability. *Journal of Education for Students Placed at Risk (JESPAR)*, *23*(1-2), 153-169.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949.

Liu, J., Kuhfeld, M., & Lee, M. (2025). Noncognitive factors and student long-run success: Comparing the predictive validity of observable academic behaviors and social-emotional skills. *Educational Policy*, *39*(1), 131-169.

Massachusetts Department of Elementary and Secondary Education. (2016). *Working with 2015 Massachusetts assessment data: Advisory from the Office of Planning and Research and the Office of Student Assessment Services*.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412-433.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences*, *2*(1), 13-43.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement, 79*(2), 310-334. https://doi.org/10.1177/0013164418783530

Succeed Boston. Boston Public Schools. (2025). https://www.bostonpublicschools.org/bps-departments/succeed-boston/our-history

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273-1296.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Wei, W. (2024). Exploring patterns of absenteeism from prekindergarten through early elementary school and their associations with children's academic outcomes. *AERA Open*, *10*(1), 1-15.

Weiland, C., Wu, T., Unterman, R., Shapiro, A., Lightner, S., Staines, T., & Taylor, A. (2025). Impacts of oversubscribed Boston Pre-K programs through middle school. EdWorkingPaper No. 25-1194. *Annenberg Institute for School Reform at Brown University.*

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838.

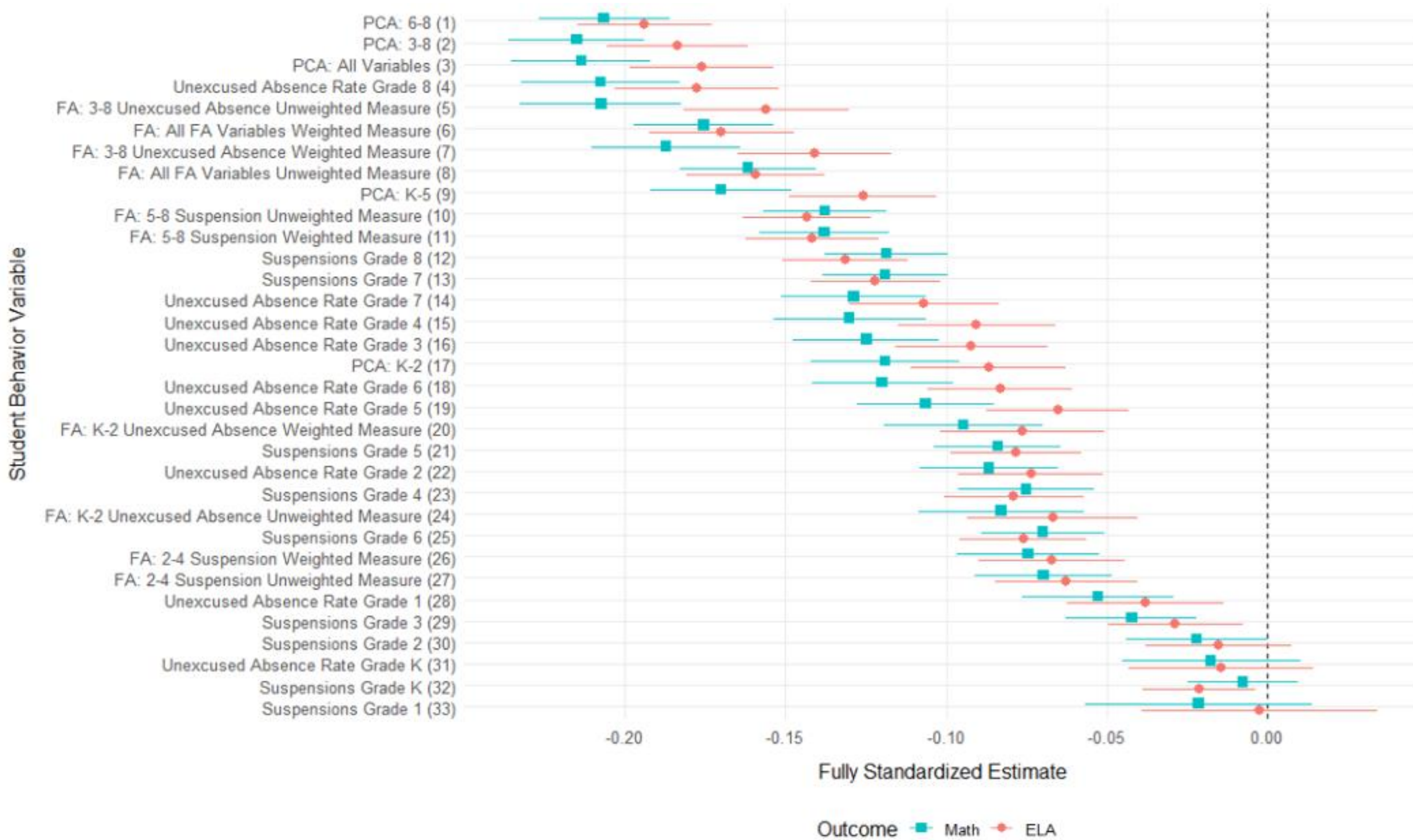
Wu, T., & Weiland, C. (2026). Leveraging modern machine learning to improve early warning systems and reduce chronic absenteeism in early childhood. *Educational Evaluation and Policy Analysis,* 1-26.

Wu, T., Weiland, C., & Staines, T. (2026). The chronic(les) of absenteeism measurement: Unpacking the many measures of attendance and evidence for a lower chronic absenteeism threshold. EdWorkingPaper No. 26-1380. *Annenberg Institute for School Reform at Brown University.*

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

Figures and Tables

Figure 1. Standardized Estimates and Confidence Intervals from Multilevel Regression Models Predicting Math and ELA Standardized Test Scores



Note: PCA = principal components analysis. FA = factor analysis. K = kindergarten. ELA = English Language Arts. All variables are fully standardized. Coefficients represent the expected change (in standard deviation units) in the outcome associated with a one standard deviation increase in the predictor, holding all else constant.

Table 1. Descriptive Statistics

	Grade									Overall Average
	Kindergarten	1	2	3	4	5	6	7	8	
Academic Behavior Indicators										
Unexcused Absence Rate (%)	1.47 (4.01)	2.01 (4.11)	2.49 (3.83)	3.02 (4.00)	3.04 (4.13)	2.96 (3.94)	3.16 (4.39)	3.46 (4.81)	4.01 (5.72)	2.83 (4.41)
Number of Suspensions	0.00 (0.09)	0.02 (0.36)	0.05 (0.53)	0.06 (0.46)	0.08 (0.50)	0.13 (0.68)	0.16 (0.70)	0.18 (0.74)	0.15 (0.59)	0.09 (0.55)
Chronic Absenteeism (%)	19.91	13.08	11.60	10.85	10.74	10.56	12.06	14.06	17.88	15.70
Ever Suspended (%)	0.23	0.72	2.05	3.18	4.31	6.68	9.03	9.99	9.23	6.11
Grade Retention (%)	0.39	2.77	4.43	3.05	1.18	0.75	0.88	1.09	1.19	1.65
Educational Attainment Outcomes										
Eighth Grade Math Standardized Test Score	--	--	--	--	--	--	--	--	--	0.233 (0.952)
Eighth Grade ELA Standardized Test Score	--	--	--	--	--	--	--	--	--	0.235 (0.933)
On-Time High School Graduation (%)	--	--	--	--	--	--	--	--	--	74.36
Immediate College Enrollment (%)	--	--	--	--	--	--	--	--	--	57.54

Note. ELA = English language arts. Unexcused absence rates and number of suspensions are reported as means with standard deviations in parentheses. Values for chronic absenteeism, ever suspended, grade retention, on-time high school graduation, and immediate college enrollment are reported as percentages (%) since they are binary indicators. For standardized test scores, we standardized each student's theta (i.e., IRT-derived) score on the mean and standard deviation of all eighth graders within Boston Public Schools taking the exam the same year, following guidance from the Massachusetts Department of Elementary & Secondary Education to facilitate longitudinal score comparisons. Missingness increased gradually across grades, ranging from 4.7%-9.9% in kindergarten, 7.1%-10.4% in first grade, 8.9%-11.6% in second grade, 10.2%-12.5% in third grade, 7.7%-13.5% in fourth grade, 8.9%-14.2% in fifth grade, 10.1%-15.7% in sixth grade, 11.5%-16.2% in seventh grade, and 13.4%-17.2% in eighth grade, reflecting student mobility and attrition over time. Eighth grade standardized tests had missingness rates of 41.6% (ELA) and 41.8% (math) due to the cancellation of statewide assessments during the COVID-19 pandemic year for the youngest cohort in our sample. On-time high school graduation and immediate college enrollment outcomes had missingness rates of 20.1% and 31.6%, respectively.

Table 2. Exploratory and Confirmatory Factor Analysis Loadings

		Model Specification						
		EFA (ML-FIML)				CFA (MLR with FIML)		
		Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Standardized Estimate**	**SE**	**R-Squared**
	Factor 1: Primary Grades Unexcused Absence Rates ("K-2 Unexcused Absence")							
	Kindergarten	**0.506**	0.084	-0.042	0.022	**0.406**	0.001	0.165
	1st grade	**0.752**	0.113	-0.062	0.056	**0.587**	0.002	0.344
	2nd grade	**0.666**	-0.105	0.049	-0.025	**0.828**	0.002	0.685
	Factor 2: Elementary and Middle Grades Unexcused Absence Rates ("3-8 Unexcused Absence")							
Unexcused Absence	3rd grade	0.407	**-0.418**	0.070	-0.072	**0.670**	0.001	0.449
	4th grade	0.244	**-0.589**	-0.005	0.002	**0.706**	0.001	0.498
	5th grade	0.099	**-0.731**	-0.019	-0.030	**0.767**	0.001	0.588
	6th grade	-0.069	**-0.882**	-0.019	-0.034	**0.711**	0.001	0.505
	7th grade	-0.111	**-0.802**	0.003	0.040	**0.690**	0.001	0.475
	8th grade	-0.055	**-0.680**	0.012	0.071	**0.609**	0.002	0.371
	Factor 3: Early Grades Suspensions ("2-4 Suspensions")							
	2nd grade	0.037	0.062	**0.570**	-0.079	**0.409**	0.010	0.167
	3rd grade	-0.009	0.004	**0.690**	-0.081	**0.570**	0.013	0.325
	4th grade	-0.043	-0.047	**0.537**	0.087	**0.604**	0.014	0.365
Suspensions	*Factor 4: Middle Grades Suspensions ("5-8 Suspensions")*							
	5th grade	0.027	0.004	0.238	**0.399**	**0.665**	0.013	0.442
	6th grade	0.023	0.011	0.084	**0.512**	**0.679**	0.014	0.461
	7th grade	0.009	0.018	-0.060	**0.756**	**0.509**	0.013	0.259
	8th grade	-0.006	-0.008	-0.084	**0.644**	**0.410**	0.010	0.168

Note: EFA = exploratory factor analysis. CFA = confirmatory factor analysis. SE = standard error. For the EFA columns, all loadings are standardized, and bolded loadings are those most strongly associated with each factor. All bolded loadings are statistically significant ($p < 0.05$). The results presented for EFA use ML-FIML, which stands for using a maximum likelihood (ML) extraction method run under full information maximum likelihood (FIML) conditions. The results for CFA use maximum likelihood estimation with robust standard errors (MLR), with missing data handled using FIML. All standardized estimates for the CFA are statistically significant ($p < 0.05$). The names shown in parentheses and quotations for each factor correspond to the labels used in the predictive validity analyses presented in Figure 1.

Table 3. Predicting On-Time High School Graduation

	1		2		3		4		5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
PCA: 6-8	-0.141***	-0.115***								
	(0.005)	(0.006)								
PCA: 3-8			-0.143***	-0.117***						
			(0.005)	(0.006)						
PCA: All Variables					-0.143***	-0.122***				
					(0.005)	(0.006)				
FA: 3-8 Unexcused Absence Unweighted Composite							-0.143***	-0.124***		
							(0.005)	(0.007)		
Unexcused Absence Rate Grade 8									-0.110***	-0.122***
									(0.005)	(0.007)
Math		0.066***		0.063***		0.062***		0.066***		0.072***
		(0.007)		(0.007)		(0.007)		(0.007)		(0.007)
ELA		0.056***		0.060***		0.061***		0.065***		0.062***
		(0.007)		(0.007)		(0.007)		(0.007)		(0.007)
Student Covariates	x	x	x	x	x	x	x	x	x	x
School Random Intercepts	x	x	x	x	x	x	x	x	x	x
N	9305	6520	9305	6520	9305	6520	9305	6520	9303	6519
AIC	9218.4	7440	9241.3	7485.2	9234.9	7526.4	9455	7635.3	9613.3	7720.7
BIC	9311.2	7546.1	9334.1	7591.3	9327.7	7632.5	9547.8	7741.4	9706.1	7826.8

Note: PCA = principal components analysis. FA = factor analysis. ELA = English language arts. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. *N* = sample size. Each column reports coefficients from a separate regression model predicting on-time high school graduation. Independent variables are standardized (mean = 0, standard deviation = 1). Coefficients represent the change in the dependent variable (in percentage points) associated with a 1 standard deviation increase in the predictor, holding all else constant. Standard errors are reported in parentheses. "PCA: 6-8" refers to the first principal component extracted from academic behavior indicators measured in grades 6-8. "PCA: 3-8" refers to the first principal component extracted from indicators measured in grades 3-8. "PCA: All Variables" refers to the first principal component extracted from all academic behavior indicators included in Table 1. "FA: 3-8 Unexcused Absence Unweighted Composite" refers to the unweighted average of the grades 3-8 unexcused absence indicators identified through the factor analytic model. "Unexcused Absence Rate Grade 8" refers to the standalone eighth-grade unexcused absence rate. Models additionally include student demographic covariates and school random intercepts. *N* varies across models because of differences in outcome availability and missing data. ***$p < 0.001$.

Table 4. Predicting Immediate College Enrollment

	1		2		3		4		5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
PCA: 6-8	-0.121***	-0.101***								
	(0.007)	(0.008)								
PCA: 3-8			-0.128***	-0.105***						
			(0.007)	(0.008)						
PCA: All Variables					-0.122***	-0.101***				
					(0.007)	(0.008)				
FA: 3-8 Unexcused Absence Unweighted Composite							-0.124***	-0.108***		
							(0.008)	(0.009)		
Unexcused Absence Rate Grade 8									-0.096***	-0.104***
									(0.007)	(0.010)
Math		0.076***		0.073***		0.074***		0.075***		0.080***
		(0.009)		(0.009)		(0.009)		(0.009)		(0.009)
ELA		0.060***		0.062***		0.063***		0.065***		0.063***
		(0.009)		(0.009)		(0.009)		(0.010)		(0.010)
Student Covariates	x	x	x	x	x	x	x	x	x	x
School Random Intercepts	x	x	x	x	x	x	x	x	x	x
N	8030	5800	8030	5800	8030	5800	8030	5800	8030	5800
AIC	10327.5	7225.1	10301.7	7217.7	10331.5	7230.5	10383.3	7236.8	10473.6	7271.2
BIC	10418.3	7325.0	10392.6	7317.7	10422.4	7330.5	10474.2	7336.8	10564.5	7371.2

Note: FA = factor analysis. ELA = English language arts. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. N = sample size. N varies across models because of differences in outcome availability and missing data. Each column reports coefficients from a separate regression model predicting immediate college enrollment. Independent variables are standardized (mean = 0, standard deviation = 1). Coefficients represent the change in the dependent variable (in percentage points) associated with a 1 standard deviation increase in the predictor, holding all else constant. Standard errors are reported in parentheses. "PCA: 6-8" refers to the first principal component extracted from academic behavior indicators measured in grades 6-8. "PCA: 3-8" refers to the first principal component extracted from indicators measured in grades 3-8. "PCA: All Variables" refers to the first principal component extracted from all academic behavior indicators included in Table 1. "FA: 3-8 Unexcused Absence Unweighted Composite" refers to the unweighted average of the grades 3-8 unexcused absence indicators identified through the factor analytic model. "Unexcused Absence Rate Grade 8" refers to the standalone eighth-grade unexcused absence rate. Models additionally include student demographic covariates and school random intercepts. PCA = principal components analysis. ***$p < 0.001$.

Appendix

Appendix A: Cohort Map

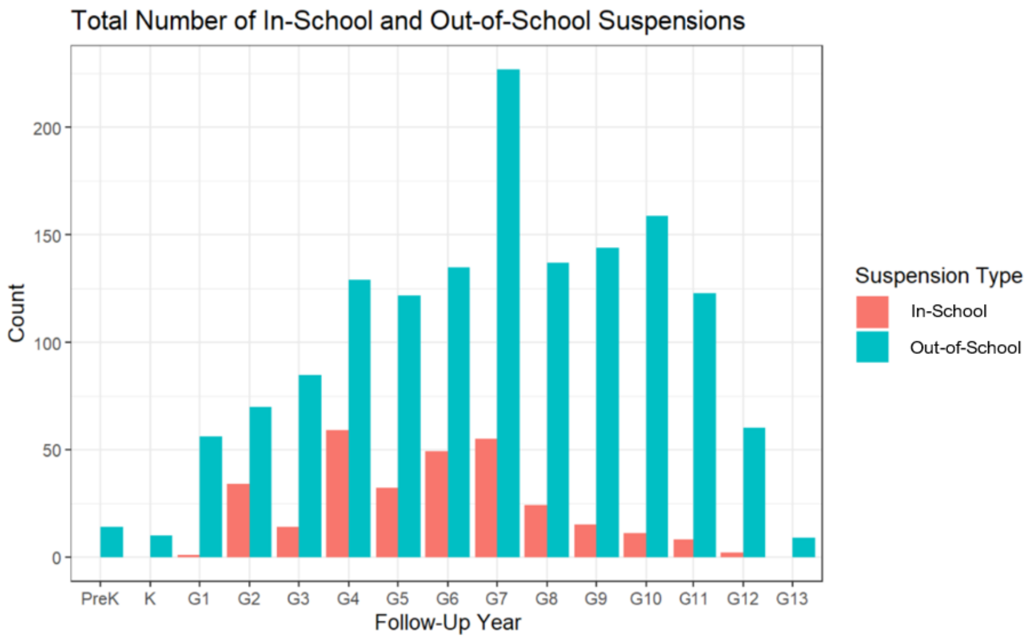
Table 1. Grade and Year Progression for Each Cohort in Sample

Cohort	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24
2007	Pre-K	K	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2008		Pre-K	K	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2009			Pre-K	K	1	2	3	4	5	6	7	8	9	10	11	12	13
2010				Pre-K	K	1	2	3	4	5	6	7	8	9	10	11	12

Note: Gray shading indicates grades used for constructing the longitudinal academic behavior composites. The vertical line between the 2019-2020 and 2020-2021 school years marks the onset of virtual (and later, hybrid instruction) school year due to the COVID-19 pandemic, which began during the spring of the 2019-2020 school year.

Appendix B: Suspensions Measurement

Figure 1. Total Number of In-School and Out-of-School Suspensions



Note: We combined in-school and out-of-school suspensions due to the structure of Boston's restorative justice intervention program, Succeed Boston, which operates similarly to an in-school suspension program but is classified as an out-of-school suspension because students are not physically present in their school buildings. Reporting different suspension types therefore conflates the traditional distinctions between in-school and out-of-school suspensions, where the latter is typically viewed as more severe. K = kindergarten. G = grade.

Appendix C: Exploratory Factor Analysis

For our EFA sample (N = 6,116), the Kaiser-Meyer-Olkin (KMO) measure of adequacy was 0.86 and Bartlett's test of sphericity was statistically significant ($p < 0.001$), indicating that the relationship between all the non-test-score variables was strong enough to conduct factor analyses (Worthington & Whittaker, 2006). To ensure robustness in identifying the factor structure, we employed three analytical approaches for the EFA (Table 1 below).

First, we used a maximum likelihood (ML) extraction method run under full information maximum likelihood conditions (FIML) within the umx package in R to address the missing data rates seen in Table 1 in the main text (Bates et al., 2019). Throughout the rest of the paper, we refer to this EFA analytical approach as ML-FIML. Unlike traditional methods such as pairwise deletion, which discard incomplete cases, FIML uses all available data to estimate parameters, enabling the inclusion of all variables without excluding cases with missing data or relying on imputation methods (Graham, 2003).

However, many variables, such as unexcused absence rate and number of suspensions, remained severely right-skewed (skewness greater than 3) even after the log transformation, violating the normality assumption underlying FIML. Therefore, we also conducted an EFA using Principal Axis Factoring (PAF) within the psych package in R, which does not require multivariate normality (Yong & Pearce, 2013). For the PAF approach, we addressed missing data through multiple imputation using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011). Both the ML-FIML and PAF approaches used the oblique (promax) rotation, as prior theoretical and empirical evidence suggested that the non-test-score factors were likely correlated (Jackson, 2018).

In these approaches, the binary variables consistently exhibited low factor loadings. We decided to exclude the binary variables from further consideration after conducting a robustness check using the Unweighted Least Squares Means and Variance (ULSMV) estimator with tetrachoric correlations to ensure the low loadings were not due to the unsuitability of these approaches for binary data (Table 2 below; Byrne, 2012; Kyriazos & Poga-Kyriazou, 2023; Watkins, 2018).

We determined the final factor solution by considering Kaiser's criterion (eigenvalues greater than one), parallel analysis (scree plot show in Figure 1 below), the model fit indices provided by different factor solutions, the internal consistency of obtained factors, and the interpretability of the obtained factor solutions. While Kaiser's criterion suggested retaining two factors, parallel analysis, a simulation-based method that compares observed eigenvalues to those from randomly generated data, recommended retaining five factors (Horn, 1965). Based on both Kaiser's criterion and the parallel analysis results, we tested two- to five-factor solutions, while also using model fit indices and interpretability to guide model selection. We excluded items that did not load onto a distinct factor of at least 3 items, along with items that did not load at 0.40 and above without significant cross-loadings onto other factors (i.e., no other loading onto another factor within 0.15 in magnitude of loading onto a factor). We determined a four-factor model to be the final EFA structure using these criteria (Worthington & Whittaker, 2006).

As seen in Table 1, both the ML-FIML and the PAF approach yielded similar loadings and the same conceptually meaningful factors reflective of time-trends in student behavior. The first factor, Primary Grades Unexcused Absence Rates, consisted of the three unexcused absence rate variables from kindergarten to second grade. The second factor, Elementary and Middle Grades Unexcused Absence Rates, consisted of the six unexcused absence rate variables from third to eighth grade. The third factor, Early Grades Suspensions, consisted of the three number of suspensions variables from second to fourth grade. The fourth factor, Middle Grades Suspensions, consisted of the four number of suspensions variables from fifth to eighth grade. These factors represented distinct patterns of student behavior over time, with unexcused absences and suspensions grouped by different developmental stages, highlighting the temporal progression and potential shifts in behavioral trends across grade levels.

Model fit indices indicated that the four-factor solution fit the data reasonably well. For the ML-FIML approach, the Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI) were 0.894 and 0.940, respectively, with the CFI exceeding the 0.90 cutoff for good fit and close to the 0.95 threshold for excellent fit (Hu & Bentler, 1998). The Root Mean Square Error of Approximation (RMSEA) was 0.053, within the 0.05–0.08 range for acceptable fit (Kline, 2011). For the PAF approach, the TLI was 0.883 (CFI was not reported in the output), the RMSEA was 0.060, and the Root Mean Square of the Residuals (RMSR) was 0.028, well below the 0.08 cutoff for good fit (Hu & Bentler, 1999).

The internal consistency of the four subscales was modest, with Cronbach's alpha for two subscales meeting the 0.7 threshold (Table 3). While the other two subscales fell between 0.6 and 0.7, prior research suggests that this range can still be considered acceptable (Taber, 2018). However, Cronbach's alpha has important limitations, particularly for shorter scales. Two of the subscales in this study include only three items, and alpha is known to underestimate reliability when item counts are low (McNeish, 2018). Therefore, we included the omega total and Coefficient H as alternative measures, which indicated higher reliability compared to Cronbach's alpha for some of the factors (McNeish, 2018). In particular, Coefficient H indicated three factors above the "acceptable" threshold of 0.7, with Coefficient H values ranging from 0.6-0.9.

When paired with the strong model fit indices, this suggests that the four-factor solution is a reasonable representation of the data.

Table 1. Exploratory Factor Analysis Loadings ($N = 6{,}116$)

		Model Specification							
Items		**ML-FIML**				**MICE imputation and PAF estimator**			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Unexcused Absence Rate	*Factor 1: Primary Grades Unexcused Absence Rates*								
	Kindergarten	**0.506**	0.084	-0.042	0.022	**0.477**	-0.05	0.013	-0.01
	1st grade	**0.752**	0.113	-0.062	0.056	**0.764**	-0.088	-0.052	0.047
	2nd grade	**0.666**	-0.105	0.049	-0.025	**0.616**	0.155	0.044	-0.015
	Factor 2: Elementary and Middle Grades Unexcused Absence Rates								
	3rd grade	0.407	**-0.418**	0.07	-0.072	0.353	**0.451**	0.04	-0.06
	4th grade	0.244	**-0.589**	-0.005	0.002	0.196	**0.607**	-0.043	0.044
	5th grade	0.099	**-0.731**	-0.019	-0.03	0.042	**0.728**	-0.064	0.022
	6th grade	-0.069	**-0.882**	-0.019	-0.034	-0.095	**0.843**	-0.026	-0.036
	7th grade	-0.111	**-0.802**	0.003	0.04	-0.134	**0.793**	0.039	-0.011
	8th grade	-0.055	**-0.68**	0.012	0.071	-0.062	**0.668**	0.031	0.04
Number of Suspensions	*Factor 3: Early Grades Suspensions*								
	2nd grade	0.037	0.062	**0.57**	-0.079	0.008	0.015	**0.63**	-0.086
	3rd grade	-0.009	0.004	**0.69**	-0.081	-0.032	0.066	**0.599**	0.005
	4th grade	-0.043	-0.047	**0.537**	0.087	-0.032	0.064	**0.461**	0.165
	Factor 4: Middle Grades Suspensions								
	5th grade	0.027	0.004	0.238	**0.399**	0.019	0.004	0.162	**0.491**
	6th grade	0.023	0.011	0.084	**0.512**	0.006	-0.001	0.018	**0.57**
	7th grade	0.009	0.018	-0.06	**0.756**	0.002	-0.018	-0.072	**0.745**
	8th grade	-0.006	-0.008	-0.084	**0.644**	-0.005	0.026	-0.08	**0.605**

Note: All loadings are standardized. Bolded loadings are those most strongly associated with each factor, and all bolded loadings are statistically significant ($p < 0.05$). ML-FIML stands for using a maximum likelihood (ML) extraction method run under full information maximum likelihood (FIML) conditions. MICE stands for Multivariate Imputation by Chained Equations. PAF stands for Principal Axis Factoring.

Table 2. Factor Loadings Using the Unweighted Least Squares Means and Variance (ULSMV) Estimator with Tetrachoric Correlations

		Factor 1	Factor 2	Factor 3
Chronic Absenteeism	Kindergarten	0.497	-0.283	0.352
	1st grade	0.631	-0.217	0.250
	2nd grade	0.692	-0.142	0.335
	3rd grade	0.784	-0.036	0.125
	4th grade	0.844	0.023	0.008
	5th grade	0.827	0.135	-0.200
	6th grade	0.801	0.165	-0.189
	7th grade	0.736	0.261	-0.163
	8th grade	0.605	0.223	-0.140
Ever Suspended	Kindergarten	-0.245	0.196	0.568
	1st grade	-0.116	0.345	0.615
	2nd grade	-0.070	0.354	0.537
	3rd grade	-0.005	0.474	0.404
	4th grade	0.051	0.531	0.325
	5th grade	-0.083	0.704	0.217
	6th grade	-0.085	0.723	0.105
	7th grade	0.004	0.721	0.048
	8th grade	-0.003	0.628	0.049
Grade Retention	Kindergarten	-0.032	-0.009	0.199
	1st grade	0.180	-0.011	0.124
	2nd grade	0.303	0.011	-0.125
	3rd grade	-0.010	-0.042	0.183
	4th grade	0.078	0.102	0.194
	5th grade	0.129	-0.050	0.317
	6th grade	0.077	0.337	-0.086
	7th grade	0.174	0.159	0.257
	8th grade	0.166	0.548	-0.207

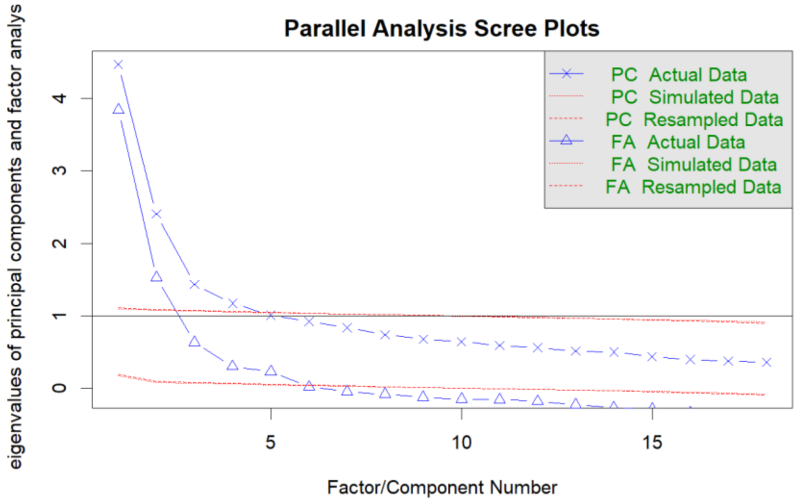
In the first two approaches, the binary variables consistently exhibited low factor loadings, likely due to the unsuitability of these approaches for binary data and the uneven distribution of each binary variable (see Table 1 for details). To address these issues, we implemented a third approach specifically tailored for binary variables, using the Unweighted Least Squares Means and Variance (ULSMV) estimator with tetrachoric correlations (Byrne, 2012). However, as seen above, this approach resulted in poor model fit statistics, with the "best-fitting" solution yielding an RMSEA of 0.4813 and a TLI of -0.098. Although negative TLI values are possible due to its non-normed nature (Shi et al., 2019), the combination of poor fit indices and a lack of interpretable factor solutions indicated that the binary variables would not meaningfully contribute to the factor structure in the EFA. As a result, we concentrate on the results of the first two approaches and exclude the binary variables from further consideration.

Table 3. Model Reliability

			Omega Total	Coefficient H	Cronbach's Alpha
EFA	Factor 1	ML-FIML	0.681	0.710	0.629
		PAF	0.656	0.698	
	Factor 2	ML-FIML	0.846	0.890	0.788
		PAF	0.843	0.874	
	Factor 3	ML-FIML	0.628	0.642	0.634
		PAF	0.584	0.598	
	Factor 4	ML-FIML	0.673	0.721	0.731
		PAF	0.698	0.724	

Note: ML-FIML stands for using a maximum likelihood (ML) extraction method run under full information maximum likelihood (FIML) conditions. PAF stands for Principal Axis Factoring. MLR stands for Maximum Likelihood Robust. MICE stands for Multivariate Imputation by Chained Equations. WLSMV stands for Weighted Least Squares Mean and Variance Adjusted.

Figure 1. Parallel Analysis Scree Plot



Appendix D: Confirmatory Factor Analysis

We cross-validated the four-factor EFA model identified in Appendix C using a CFA, with an independent sample of randomly selected participants ($N = 6{,}116$; Worthington & Whittaker, 2006). Because we used an oblique rotation in the EFA, we also estimated correlations between all factors in the CFA, aligning with the theoretical expectation that "noncognitive" behaviors are interrelated. Additionally, we correlated residual terms of repeated observed items for both the unexcused absence and suspension factors. This adjustment accounts for the temporal dependencies between repeated measures of the same item administered to the same student over time, which may otherwise inflate the error variance (Furr, 2021). To ensure robustness, we employed three analytical approaches using the *lavaan* package in R.

First, we used the maximum likelihood estimation with robust standard errors (MLR). This estimator is well-suited for data that may deviate from normality, as it adjusts standard errors to account for non-normal distributions (Rosseel, 2012). Missing data were handled using full information maximum likelihood (FIML). Second, we conducted another CFA using MLR but using a dataset where missing values were imputed through MICE to provide a robustness check over the FIML missing data handling method. Lastly, we conducted a CFA using the Weighted Least Squares with Mean and Variance adjustment (WLSMV) estimator also using MICE to impute missing values. WLSMV is particularly well-suited for ordinal or count-like data, as it does not assume multivariate normality, making it ideal for variables with a limited range of discrete values like the number of suspensions. By treating these variables as ordinal, WLSMV could allow for a more appropriate handling of their discrete nature (Li, 2016).

Table 1 below provides standardized factor loadings, standard errors, and R-squared values for the three approaches. All standardized loadings were statistically significant ($p < 0.05$)

and above 0.4 in magnitude. The pattern of association among factors in the CFA was similar to the pattern in the EFA, supporting the correlation between different "noncognitive" skill domains (Table 2), and each approach produced similar model fit indices (Table 3).

To evaluate model fit, we relied on the robust versions of the CFI, TLI, and the Root Mean Square Error of Approximation (RMSEA), as these metrics are well-suited for data that deviate from multivariate normality (Bentler, 1990; Hu & Bentler, 1999). All robust CFI and TLI metrics were above the 0.95 cutoffs proposed in the literature, RMSEA was all under 0.05, and the SRMR was approximately 0.03 for all approaches. In sum, these model fit indices support the factor structure identified via EFA and replicated with the CFA approach. The internal consistency (Table 4) remained still modest and similar to values observed in the EFA. One explanation for this could be that administrative indicators for suspensions and absences may not be consistent over time because of its low incidence, leading to lower internal consistency. Despite this, the strong model fit indices and the replication of factor loadings from the EFA to the CFA suggest that the factor structure is robust and meaningful.

Table 1. Confirmatory Factor Analysis Loadings (*N* = 6,116)

Items		Model Specification								
		MLR with FIML			MLR with MICE Imputation			WLSMV with MICE Imputation		
		Standardized Estimate	SE	R-Squared	Standardized Estimate	SE	R-Squared	Standardized Estimate	SE	R-Squared
Unexcused Absence Rate	*Factor 1: Primary Grades Unexcused Absence Rates*									
	Kindergarten	0.406	0.001	0.165	0.421	0.001	0.177	0.424	0.001	0.180
	1st grade	0.587	0.002	0.344	0.611	0.002	0.374	0.634	0.002	0.401
	2nd grade	0.828	0.002	0.685	0.843	0.001	0.711	0.870	0.002	0.757
	Factor 2: Elementary and Middle Grades Unexcused Absence Rates									
	3rd grade	0.670	0.001	0.449	0.647	0.001	0.418	0.675	0.001	0.455
	4th grade	0.706	0.001	0.498	0.676	0.001	0.456	0.685	0.001	0.469
	5th grade	0.767	0.001	0.588	0.764	0.001	0.583	0.693	0.001	0.480
	6th grade	0.711	0.001	0.505	0.719	0.001	0.517	0.640	0.001	0.410
	7th grade	0.690	0.001	0.475	0.696	0.001	0.484	0.644	0.001	0.415
	8th grade	0.609	0.002	0.371	0.607	0.002	0.369	0.594	0.002	0.352
Number of Suspensions	*Factor 3: Early Grades Suspensions*									
	2nd grade	0.409	0.010	0.167	0.412	0.010	0.170	0.407	0.009	0.166
	3rd grade	0.570	0.013	0.325	0.577	0.013	0.333	0.637	0.015	0.406
	4th grade	0.604	0.014	0.365	0.603	0.013	0.364	0.612	0.014	0.375
	Factor 4: Middle Grades Suspensions									
	5th grade	0.665	0.013	0.442	0.657	0.012	0.432	0.634	0.012	0.402
	6th grade	0.679	0.014	0.461	0.688	0.013	0.474	0.584	0.014	0.341
	7th grade	0.509	0.013	0.259	0.521	0.012	0.272	0.545	0.013	0.297
	8th grade	0.410	0.01	0.168	0.413	0.009	0.170	0.443	0.01	0.196

Note: FIML stands for Full Information Maximum Likelihood. MLR stands for Maximum Likelihood Robust. MICE stands for Multivariate Imputation by Chained Equations. WLSMV stands for Weighted Least Squares Mean and Variance Adjusted. SE stands for standard error. All standardized estimates are statistically significant ($p < 0.05$).

Table 2. Factor Correlations for EFA and CFA

		Factor 1: Primary Grades Unexcused Absence Rates	Factor 2: Elementary and Middle Grades Unexcused Absence Rates	Factor 3: Early Grades Suspensions	Factor 4: Middle Grades Suspensions
EFA: FIML	Factor 1	1.00			
	Factor 2	-0.578	1.00		
	Factor 3	0.246	-0.197	1.00	
	Factor 4	0.141	-0.263	0.533	1.00
EFA: PAF	Factor 1	1.00			
	Factor 2	0.543	1.00		
	Factor 3	0.276	0.141	1.00	
	Factor 4	0.117	0.209	0.489	1.00
CFA: MLR with FIML	Factor 1	1.00			
	Factor 2	0.655	1.00		
	Factor 3	0.207	0.226	1.00	
	Factor 4	0.124	0.282	0.728	1.00
CFA: MLR with MICE	Factor 1	1.00			
	Factor 2	0.643	1.00		
	Factor 3	0.202	0.221	1.00	
	Factor 4	0.132	0.293	0.712	1.00
CFA: WLSMV	Factor 1	1.00			
	Factor 2	0.648	1.00		
	Factor 3	0.201	0.213	1.00	
	Factor 4	0.132	0.338	0.644	1.00

Note: FIML stands for Full Information Maximum Likelihood. PAF stands for Principal Axis Factoring. MLR stands for Maximum Likelihood Robust. MICE stands for Multivariate Imputation by Chained Equations. WLSMV stands for Weighted Least Squares Mean and Variance Adjusted.

Table 3. Confirmatory Factor Analysis Model Fit

	Model Specification		
	MLR with FIML	**MLR with MICE**	**WLSMV with MICE**
Robust CFI	0.971	0.966	0.984
Robust TLI	0.958	0.951	0.977
Robust RMSEA	0.041	0.044	0.017
SRMR	0.029	0.031	0.030

Note: FIML stands for Full Information Maximum Likelihood. MLR stands for Maximum Likelihood Robust. MICE stands for Multivariate Imputation by Chained Equations. WLSMV stands for Weighted Least Squares Mean and Variance Adjusted. SE stands for standard error.

Table 4. Model Reliability

			Omega Total	Coefficient H	Cronbach's Alpha
CFA	Factor 1	MLR with FIML	0.648	0.744	0.622
		MLR with MICE	0.669	0.766	
		WLSMV with MICE	0.691	0.800	
	Factor 2	MLR with FIML	0.847	0.852	0.805
		MLR with MICE	0.842	0.847	
		WLSMV with MICE	0.819	0.821	
	Factor 3	MLR with FIML	0.539	0.557	0.609
		MLR with MICE	0.543	0.560	
		WLSMV with MICE	0.572	0.597	
	Factor 4	MLR with FIML	0.657	0.688	0.724
		MLR with MICE	0.662	0.691	
		WLSMV with MICE	0.638	0.650	

Note: ML-FIML stands for using a maximum likelihood (ML) extraction method run under full information maximum likelihood (FIML) conditions. PAF stands for Principal Axis Factoring. MLR stands for Maximum Likelihood Robust. MICE stands for Multivariate Imputation by Chained Equations. WLSMV stands for Weighted Least Squares Mean and Variance Adjusted.

Appendix E: Principal Components Analysis

We created five PCA-based indices based on measures used by Gray-Lobe et al. (2023) and Jackson (2018). Gray-Lobe et al. (2023) developed a disciplinary index by extracting the first principal component of several high school disciplinary measures (e.g., ever suspended, number of suspensions, truancy measures, days absent, and juvenile incarceration from grades 9-12). Similarly, Jackson (2018) constructed a behavior index using the first principal component of ninth-grade measures, including GPA, grade retention, absences, and suspension status. We created one PCA index using the first principal component of all disciplinary measures included in our EFA models, as listed in Table 1 of the main text. Additionally, we generated four grade-specific PCA indices to account for potential temporal trends as observed in the factor analysis: one just using measures from grades K-2, one using measures from elementary and middle grades 3-8, one using measures from grades K-5, and one using measures, and one using measures from grades 6-8. The first two grade groupings were informed by the grade splits identified in the factor analysis for unexcused absences. The latter two groupings reflect conceptual distinctions commonly made between early and later schooling stages.

Table 1 below summarizes the five indicators with the largest loadings for each PCA specification, along with the proportion of total variance explained by the first principal component. Across all specifications, unexcused absence rates and chronic absenteeism measures consistently exhibited the largest loadings, indicating that absenteeism contributed more strongly to the first principal component than suspension-related indicators.

The PCA results also suggest that students' academic behaviors are not well represented by a single underlying dimension. The scree plot of the PCA with all variables (Figure 1; findings were similar for the other PCA results as seen in Figures 2-5) indicated multiple dimensions with an eigenvalue greater than one and a steep drop in eigenvalues after the first three components, suggesting that three components should be extracted. The first principal component explained 16.3% of the variance, while the first three components jointly explained 30.4% of the variance. Thus, although the first principal component captured the dominant dimension of variation, it represented only a modest share of the overall variability in the administrative indicators.

Because our scree plots indicated more than one component with an eigenvalue greater than one for each PCA specification (see Figure 1), we also conducted a robustness check examining whether the second principal component (PC2) of each specification held comparable predictive validity. We re-estimated our eighth grade ELA and math test score models (Research Question 3), replacing each measure based on the first principal component (PC1) with its corresponding PC2-based measure.

As shown in Table 2, PC2-based measures were considerably weaker in magnitudeand less consistent predictors of eighth grade test scores than their PC1 counterparts reported in the main text (Main Text Figure 1; Appendix F, Table 1). Standardized estimates for PC2 ranged from -0.077 to 0.083 for ELA and -0.053 to 0.060 for math. Notably, the PC2 measure for grades 6-8 was positively associated with both ELA and math scores, the opposite direction from the negative associations observed for every PC1-based and factor-analytic measure in the main analyses. The PCA: 3-8 and PCA: K-5 PC2 measures were also not statistically significant for at least one outcome.

These results suggest that PC2 does not capture a substantively meaningful or consistently interpretable dimension of academic behavior. This is consistent with our finding that the first principal component in each specification was heavily and consistently loaded on unexcused absence and chronic absenteeism measures (Table 1), while later components likely reflect a mix of residual variance without a clear behavioral interpretation. Given the weaker, less consistent, and in some cases sign-reversed associations of PC2 with our outcomes of interest, we used the first principal component as our primary PCA-based measure in the main text, consistent with prior literature using single-component behavioral indices (Gray-Lobe et al., 2023; Jackson, 2018). This choice facilitates direct comparisons with the existing literature while also highlighting a limitation of the PCA approach: a single principal component may provide only a partial summary of the multidimensional patterns of student academic behavior based on the administrative indicators we used.

Table 1. Top Five Indicator Loadings and Variance Explained by the First Principal Component Across Five PCA Specifications

	Top 5 Indicators	Loading	Total Variance Explained for First Principal Component
PCA: All Variables	Unexcused Absence Rate Grade 5	0.244	0.163
	Unexcused Absence Rate Grade 4	0.243	
	Unexcused Absence Rate Grade 7	0.241	
	Unexcused Absence Rate Grade 3	0.241	
	Unexcused Absence Rate Grade 6	0.240	
PCA: 6-8	Unexcused Absence Rate Grade 7	0.336	0.245
	Unexcused Absence Rate Grade 8	0.330	
	Chronic Absenteeism Grade 7	0.327	
	Chronic Absenteeism Grade 8	0.317	
	Unexcused Absence Rate Grade 6	0.305	
PCA: 3-8	Unexcused Absence Rate Grade 7	0.270	0.202
	Unexcused Absence Rate Grade 6	0.265	
	Unexcused Absence Rate Grade 5	0.263	
	Unexcused Absence Rate Grade 4	0.256	
	Chronic Absenteeism Grade 7	0.255	
PCA: K-5	Unexcused Absence Rate Grade 3	0.311	0.179
	Unexcused Absence Rate Grade 4	0.299	
	Chronic Absenteeism Grade 3	0.287	
	Unexcused Absence Rate Grade 2	0.285	
	Unexcused Absence Rate Grade 5	0.281	
PCA: K-2	Unexcused Absence Rate Grade 2	0.388	0.195
	Chronic Absenteeism Grade 1	0.369	
	Unexcused Absence Rate Grade 1	0.363	
	Chronic Absenteeism Grade 2	0.362	
	Chronic Absenteeism Grade K	0.320	

Table 2. Standardized Estimates from Multilevel Regression Models Predicting Standardized ELA and Math Test Scores Using the Second Principal Component (PC2) of Each PCA Specification

Variable	Standardized Estimate	Unstandardized Estimate	SE	t-statistic
ELA				
PCA: 6-8	0.083***	0.038	0.005	7.745
PCA: 3-8	-0.016	-0.010	0.007	-1.358
PCA: All Variables	-0.077***	-0.038	0.005	-7.444
PCA: K-5	-0.034**	-0.017	0.006	-2.993
PCA: K-2	-0.062***	-0.036	0.006	-5.898
Math				
PCA: 6-8	0.060***	0.028	0.005	5.757
PCA: 3-8	-0.038**	-0.023	0.007	-3.323
PCA: All Variables	-0.053***	-0.026	0.005	-5.252
PCA: K-5	-0.017	-0.009	0.006	-1.507
PCA: K-2	-0.037***	-0.022	0.006	-3.668

Note: SE = standard error. ***$p < 0.001$, **$p < 0.01$.

Figure 1. Scree Plot of PCA with All Variables

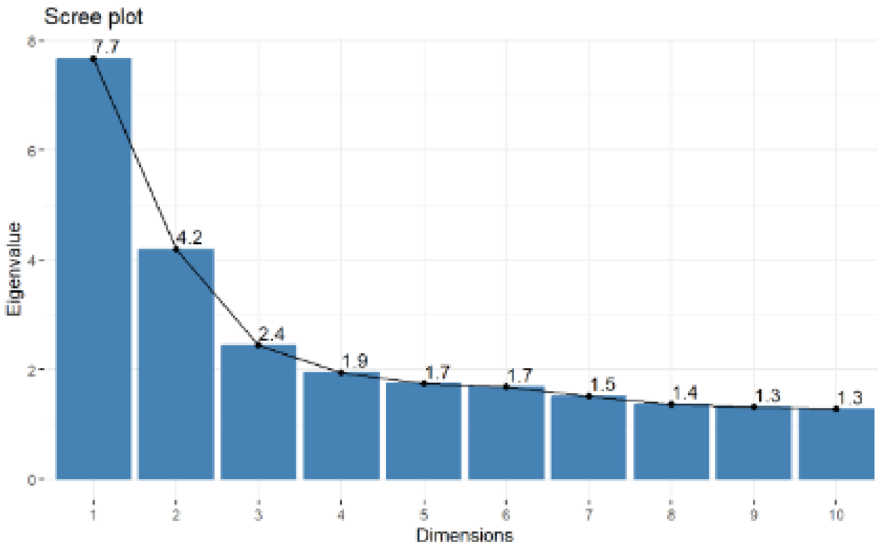


Figure 2. Scree Plot of PCA for Grades K-2

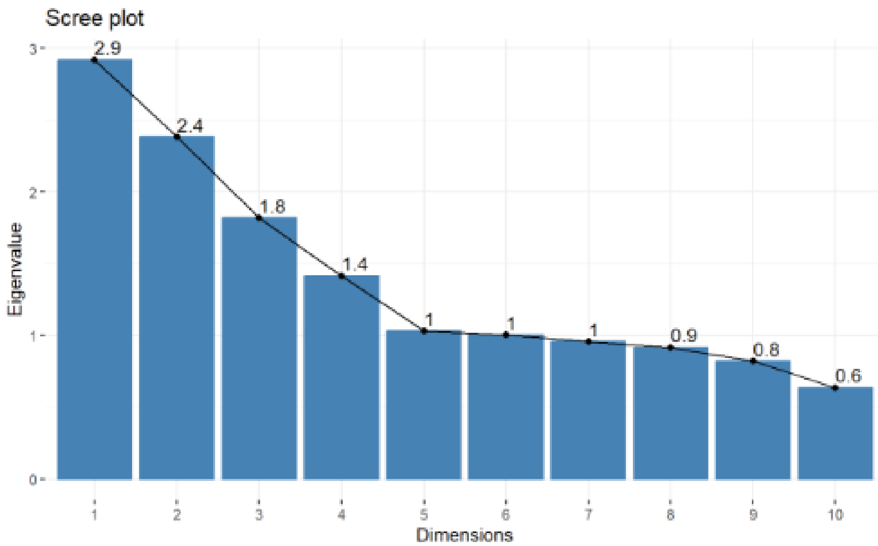


Figure 3. Scree Plot of PCA for Grades 3-8

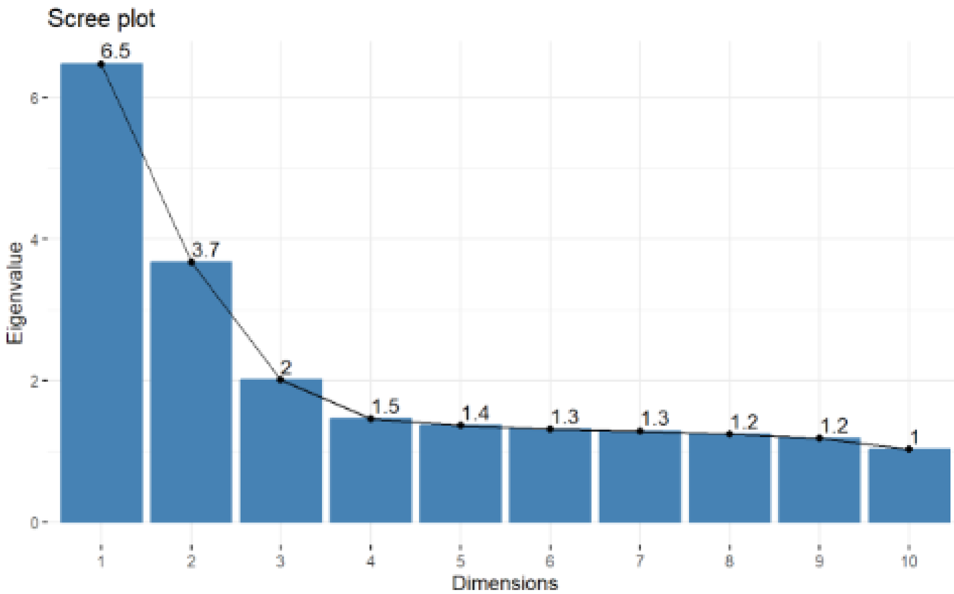


Figure 4. Scree Plot of PCA for Grade K-5

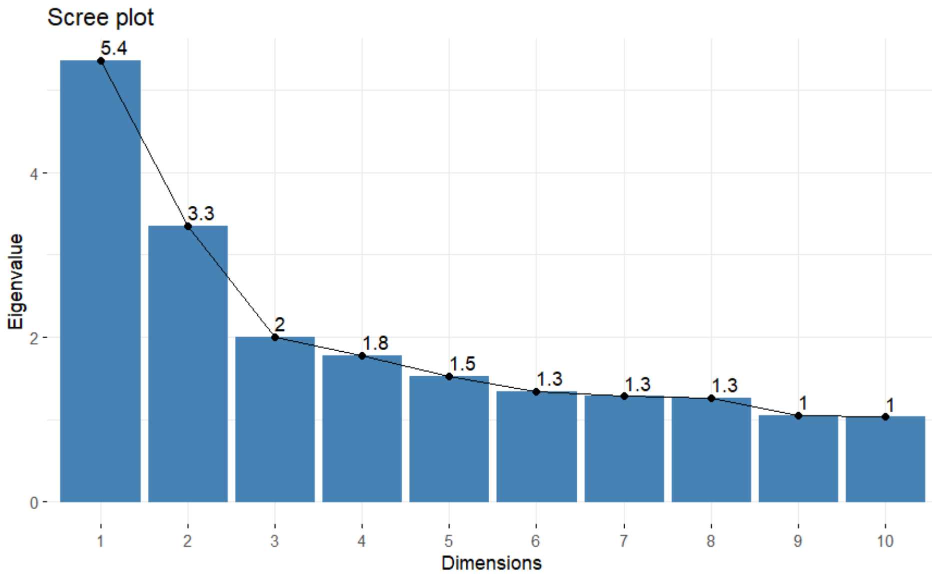
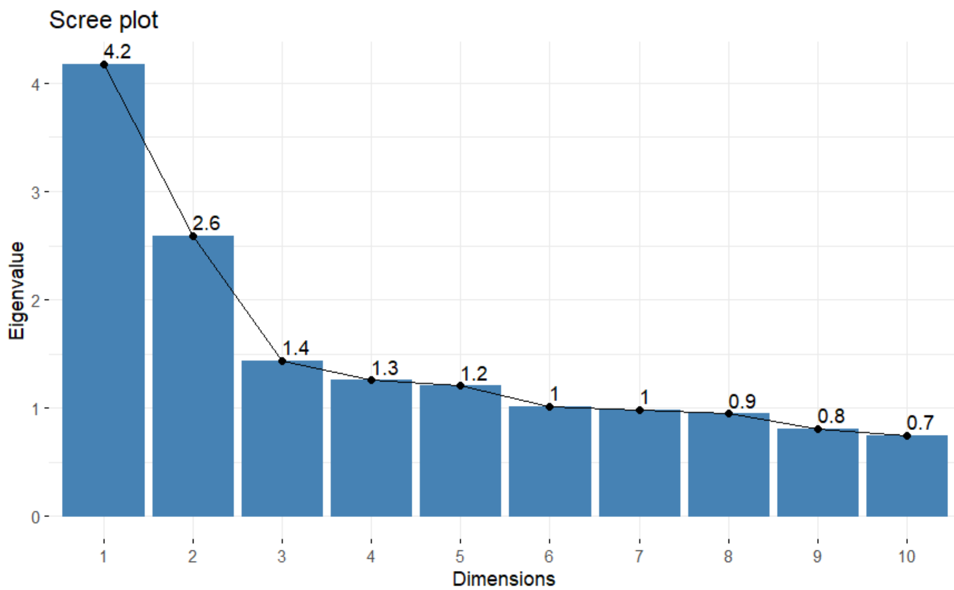


Figure 5. Scree Plot of PCA for Grade 6-8



Appendix F: Predictive Validity for Eighth Grade Standardized Testing

Table 1. Standardized Estimates from Multilevel Regression Models Predicting Standardized ELA Test Scores using School Random Intercepts

Variable	Standardized Estimate	Unstandardized Estimate	SE	t-statistic
Factor Analysis (FA)				
All FA Variables Weighted	-0.170***	-0.956	0.065	-14.746
All FA Variables Unweighted	-0.159***	-1.944	0.135	-14.377
3-8 Unexcused Absence Unweighted	-0.156***	-4.344	0.368	-11.814
5-8 Suspension Unweighted	-0.143***	-0.658	0.046	-14.182
5-8 Suspension Weighted	-0.142***	-0.275	0.020	-13.438
3-8 Unexcused Absence Weighted	-0.141***	-1.068	0.093	-11.547
K-2 Unexcused Absence Weighted	-0.076***	-1.409	0.242	-5.814
2-4 Suspension Weighted	-0.067***	-0.277	0.048	-5.777
K-2 Unexcused Absence Unweighted	-0.067***	-2.141	0.435	-4.916
2-4 Suspension Unweighted	-0.063***	-0.432	0.078	-5.518
Principal Components Analysis (PCA)				
PCA: 6-8	-0.194***	-0.088	0.005	-18.126
PCA: 3-8	-0.184***	-0.067	0.004	-16.418
PCA: All Variables	-0.176***	-0.059	0.004	-15.305
PCA: K-5	-0.126***	-0.051	0.005	-10.742
PCA: K-2	-0.087***	-0.047	0.007	-7.046
Standalone Point-in-Time Indicators				
Unexcused Absence Rate Grade 8	-0.178***	-3.278	0.241	-13.620
Suspensions Grade 8	-0.131***	-0.448	0.034	-13.148
Suspensions Grade 7	-0.122***	-0.370	0.031	-11.918
Unexcused Absence Rate Grade 7	-0.107***	-2.316	0.257	-8.997
Unexcused Absence Rate Grade 3	-0.092***	-2.369	0.310	-7.630
Unexcused Absence Rate Grade 4	-0.091***	-2.285	0.317	-7.217
Unexcused Absence Rate Grade 6	-0.083***	-1.968	0.272	-7.228
Suspensions Grade 4	-0.079***	-0.358	0.050	-7.114
Suspensions Grade 5	-0.078***	-0.277	0.037	-7.529
Suspensions Grade 6	-0.076***	-0.241	0.032	-7.524
Unexcused Absence Rate Grade 2	-0.074***	-1.944	0.302	-6.437
Unexcused Absence Rate Grade 5	-0.065***	-1.688	0.294	-5.752
Unexcused Absence Rate Grade 1	-0.038**	-0.962	0.317	-3.034
Suspensions Grade 3	-0.029**	-0.147	0.055	-2.678
Suspensions Grade K	-0.021*	-0.419	0.177	-2.365
Suspensions Grade 2	-0.015	-0.085	0.065	-1.299
Unexcused Absence Rate Grade K	-0.014	-0.368	0.378	-0.975
Suspensions Grade 1	-0.002	-0.021	0.169	-0.127

Note: Variables are ordered by descending standardized estimate magnitude within each approach. SE = standard error. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

Table 2. Standardized Estimates from Multilevel Regression Models Predicting Standardized Math Test Scores using School Random Intercepts

Variable	Standardized Estimate	Unstandardized Estimate	SE	t-statistic
Factor Analysis (FA)				
3-8 Unexcused Absence Unweighted	-0.208***	-5.916	0.366	-16.159
3-8 Unexcused Absence Weighted	-0.187***	-1.454	0.091	-15.898
All FA Variables Weighted	-0.175***	-1.011	0.064	-15.713
All FA Variables Unweighted	-0.162***	-2.022	0.134	-15.042
5-8 Suspension Weighted	-0.138***	-0.274	0.020	-13.436
5-8 Suspension Unweighted	-0.138***	-0.647	0.046	-14.014
K-2 Unexcused Absence Weighted	-0.095***	-1.789	0.239	-7.472
K-2 Unexcused Absence Unweighted	-0.083***	-2.713	0.431	-6.301
2-4 Suspension Weighted	-0.075***	-0.315	0.048	-6.612
2-4 Suspension Unweighted	-0.070***	-0.492	0.078	-6.351
Principal Components Analysis (PCA)				
PCA: 3-8	-0.215***	-0.081	0.004	-19.897
PCA: All Variables	-0.214***	-0.074	0.004	-19.263
PCA: 6-8	-0.207***	-0.097	0.005	-19.917
PCA: K-5	-0.170***	-0.070	0.005	-15.096
PCA: K-2	-0.119***	-0.066	0.007	-10.048
Standalone Point-in-Time Indicators				
Unexcused Absence Rate Grade 8	-0.208***	-3.923	0.240	-16.379
Unexcused Absence Rate Grade 4	-0.130***	-3.359	0.312	-10.753
Unexcused Absence Rate Grade 7	-0.129***	-2.856	0.255	-11.192
Unexcused Absence Rate Grade 3	-0.125***	-3.286	0.306	-10.744
Unexcused Absence Rate Grade 6	-0.120***	-2.905	0.271	-10.705
Suspensions Grade 7	-0.119***	-0.370	0.031	-12.002
Suspensions Grade 8	-0.119***	-0.415	0.034	-12.22
Unexcused Absence Rate Grade 5	-0.106***	-2.825	0.290	-9.740
Unexcused Absence Rate Grade 2	-0.087***	-2.341	0.299	-7.827
Suspensions Grade 5	-0.084***	-0.305	0.037	-8.341
Suspensions Grade 4	-0.075***	-0.348	0.050	-6.975
Suspensions Grade 6	-0.070***	-0.227	0.032	-7.119
Unexcused Absence Rate Grade 1	-0.053***	-1.370	0.313	-4.371
Suspensions Grade 3	-0.042***	-0.222	0.055	-4.071
Suspensions Grade 2	-0.022*	-0.127	0.065	-1.968
Suspensions Grade 1	-0.021	-0.198	0.167	-1.185
Unexcused Absence Rate Grade K	-0.017	-0.462	0.375	-1.234
Suspensions Grade K	-0.007	-0.152	0.178	-0.854

Note: Variables are ordered by descending standardized estimate magnitude within each approach. SE = standard error. ***$p < 0.001$, *$p < 0.05$.

Table 3. Standardized Estimates from Fixed Effects Regression Models Predicting Standardized ELA Test Scores

Variable	Standardized Estimate	Unstandardized Estimate	SE	t-statistic
Factor Analysis (FA)				
FA: All FA Variables Weighted	-0.159***	-0.893	0.067	-13.41
FA: All FA Variables Unweighted	-0.151***	-1.838	0.138	-13.299
FA: 3-8 Unexcused Absence Unweighted	-0.139***	-3.876	0.382	-10.136
FA: 5-8 Suspension Unweighted	-0.138***	-0.632	0.047	-13.339
FA: 5-8 Suspension Weighted	-0.136***	-0.263	0.021	-12.601
FA: 3-8 Unexcused Absence Weighted	-0.124***	-0.943	0.096	-9.786
FA: K-2 Unexcused Absence Weighted	-0.067***	-1.24	0.246	-5.033
FA: 2-4 Suspension Weighted	-0.062***	-0.256	0.049	-5.258
FA: K-2 Unexcused Absence Unweighted	-0.059***	-1.877	0.441	-4.253
FA: 2-4 Suspension Unweighted	-0.058***	-0.400	0.079	-5.050
Principal Components Analysis (PCA)				
PCA: 6-8	-0.183***	-0.084	0.005	-16.657
PCA: 3-8	-0.17***	-0.062	0.004	-14.772
PCA: All Variables	-0.162***	-0.054	0.004	-13.637
PCA: K-5	-0.112***	-0.045	0.005	-9.314
PCA: K-2	-0.076***	-0.041	0.007	-6.082
Standalone Point-in-Time Indicators				
Unexcused Absence Rate Grade 8	-0.166***	-3.067	0.248	-12.362
Suspensions Grade 8	-0.128***	-0.438	0.035	-12.527
Suspensions Grade 7	-0.116***	-0.352	0.032	-11.142
Unexcused Absence Rate Grade 7	-0.096***	-2.075	0.267	-7.779
Unexcused Absence Rate Grade 3	-0.079***	-2.02	0.318	-6.353
Unexcused Absence Rate Grade 4	-0.077***	-1.949	0.327	-5.959
Suspensions Grade 5	-0.074***	-0.262	0.037	-7.027
Suspensions Grade 4	-0.073***	-0.331	0.051	-6.485
Unexcused Absence Rate Grade 6	-0.073***	-1.725	0.28	-6.154
Suspensions Grade 6	-0.071***	-0.226	0.032	-6.969
Unexcused Absence Rate Grade 2	-0.066***	-1.736	0.307	-5.655
Unexcused Absence Rate Grade 5	-0.054***	-1.392	0.305	-4.557
Unexcused Absence Rate Grade 1	-0.031*	-0.798	0.322	-2.48
Suspensions Grade 3	-0.027*	-0.139	0.056	-2.474
Suspensions Grade K	-0.020*	-0.395	0.183	-2.165
Suspensions Grade 2	-0.014	-0.078	0.066	-1.179
Unexcused Absence Rate Grade K	-0.013	-0.327	0.382	-0.857
Suspensions Grade 1	0.001	0.009	0.171	0.055

Note: Variables are ordered by descending standardized estimate magnitude within each approach. SE = standard error. ***$p < 0.001$, *$p < 0.05$.

Table 4. Standardized Estimates from Fixed Effects Regression Models Predicting Standardized Math Test Scores

Variable	Standardized Estimate	Unstandardized Estimate	SE	t-statistic
Factor Analysis (FA)				
FA: 3-8 Unexcused Absence Unweighted	-0.197***	-5.604	0.378	-14.824
FA: 3-8 Unexcused Absence Weighted	-0.177***	-1.374	0.094	-14.544
FA: All FA Variables Weighted	-0.167***	-0.963	0.066	-14.677
FA: All FA Variables Unweighted	-0.154***	-1.927	0.137	-14.094
FA: 5-8 Suspension Unweighted	-0.133***	-0.623	0.047	-13.257
FA: 5-8 Suspension Weighted	-0.132***	-0.263	0.021	-12.686
FA: K-2 Unexcused Absence Weighted	-0.088***	-1.664	0.243	-6.857
FA: K-2 Unexcused Absence Unweighted	-0.077***	-2.512	0.435	-5.768
FA: 2-4 Suspension Weighted	-0.070***	-0.296	0.048	-6.137
FA: 2-4 Suspension Unweighted	-0.065***	-0.461	0.078	-5.879
Principal Components Analysis (PCA)				
PCA: 3-8	-0.206***	-0.077	0.004	-18.569
PCA: All Variables	-0.204***	-0.070	0.004	-17.941
PCA: 6-8	-0.198***	-0.092	0.005	-18.601
PCA: K-5	-0.161***	-0.067	0.005	-14.061
PCA: K-2	-0.112***	-0.062	0.007	-9.309
Standalone Point-in-Time Indicators				
Unexcused Absence Rate Grade 8	-0.199***	-3.756	0.247	-15.216
Unexcused Absence Rate Grade 4	-0.124***	-3.212	0.321	-10.013
Unexcused Absence Rate Grade 7	-0.120***	-2.657	0.263	-10.086
Unexcused Absence Rate Grade 3	-0.117***	-3.089	0.312	-9.898
Suspensions Grade 8	-0.116***	-0.404	0.035	-11.653
Suspensions Grade 7	-0.114***	-0.355	0.031	-11.356
Unexcused Absence Rate Grade 6	-0.113***	-2.730	0.278	-9.833
Unexcused Absence Rate Grade 5	-0.099***	-2.639	0.300	-8.796
Suspensions Grade 5	-0.080***	-0.291	0.037	-7.877
Unexcused Absence Rate Grade 2	-0.080***	-2.174	0.303	-7.176
Suspensions Grade 4	-0.071***	-0.331	0.051	-6.557
Suspensions Grade 6	-0.066***	-0.213	0.032	-6.616
Unexcused Absence Rate Grade 1	-0.049***	-1.277	0.317	-4.024
Suspensions Grade 3	-0.040***	-0.212	0.056	-3.822
Suspensions Grade 1	-0.019	-0.180	0.169	-1.062
Suspensions Grade 2	-0.019	-0.107	0.065	-1.650
Unexcused Absence Rate Grade K	-0.015	-0.408	0.377	-1.083
Suspensions Grade K	-0.007	-0.142	0.183	-0.780

Note: Variables are ordered by descending standardized estimate magnitude within each approach. SE = standard error. ***$p < 0.001$.

Appendix G: Predictive Validity for High School Graduation and College Enrollment

Table 1. Predicting On-Time High School Graduation with 7th Grade Standardized Test Scores

	1		2		3		4		5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
PCA: 6-8	-0.141***	-0.121***								
	(0.005)	(0.005)								
PCA: 3-8			-0.143***	-0.120***						
			(0.005)	(0.005)						
PCA: All Variables					-0.143***	-0.116***				
					(0.005)	(0.005)				
FA: 3-8 Unexcused Absence Unweighted Composite							-0.143***	-0.118***		
							(0.005)	(0.005)		
Unexcused Absence Rate Grade 8									-0.110***	-0.096***
									(0.005)	(0.005)
7th Grade Math		0.061***		0.058***		0.058***		0.063***		0.058***
		(0.007)		(0.007)		(0.007)		(0.007)		(0.007)
7th Grade ELA		0.045***		0.049***		0.051***		0.054***		0.054***
		(0.006)		(0.006)		(0.006)		(0.006)		(0.006)
Student Covariates	x	x	x	x	x	x	x	x	x	x
School Random Intercepts	x	x	x	x	x	x	x	x	x	x
N	9305	8721	9305	8721	9305	8721	9305	8721	9303	8720
AIC	9218.4	7440	9241.3	7485.2	9234.9	7526.4	9455	7635.3	9613.3	7720.7
BIC	9311.2	7546.1	9334.1	7591.3	9327.7	7632.5	9547.8	7741.4	9706.1	7826.8

Table 2. Predicting Immediate College Attendance with 7th Grade Standardized Test Scores

	1		2		3		4		5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
PCA: 6-8	-0.121***	-0.100***								
	(0.007)	(0.007)								
PCA: 3-8			-0.128***	-0.105***						
			(0.007)	(0.007)						
PCA: All Variables					-0.122***	-0.099***				
					(0.007)	(0.007)				
FA: 3-8 Unexcused Absence Unweighted Composite							-0.124***	-0.103***		
							(0.008)	(0.008)		
Unexcused Absence Rate Grade 8									-0.096***	-0.081***
									(0.007)	(0.016)
7th Grade Math		0.080***		0.077***		0.078***		0.080***		0.085***
		(0.008)		(0.008)		(0.008)		(0.008)		(0.008)
7th Grade ELA		0.049***		0.051***		0.052***		0.053***		0.053***
		(0.008)		(0.008)		(0.008)		(0.009)		(0.009)
Student Covariates	x	x	x	x	x	x	x	x	x	x
School Random Intercepts	x	x	x	x	x	x	x	x	x	x
N	8030	7847	8030	7847	8030	7847	8030	7847	8030	7847
AIC	10327.5	9798.7	10301.7	9779.6	10331.5	9799.8	10383.3	9822.3	10473.6	9888.6
BIC	10418.3	9903.2	10392.6	9884.1	10422.4	9904.3	10474.2	9926.8	10564.5	9993.1