# YouthTruth Concurrent and Predictive Validity Study

August 2022


**Prepared for:**
YouthTruth
131 Steuart Street, Suite 501
San Francisco, CA 94105

**Prepared by:**
SRI International
Daniel Princiotta, PhD
Kyra Caspary, PhD


SRI Education™
A DIVISION OF SRI INTERNATIONAL

**Acknowledgements:**
We are grateful to the members of the YouthTruth Research Advisory Committee—Elisha Smith-Arrilaga, Ellie Buteau, and John Easton—who were generous in contributing their time and expertise to shape and guide this study. Further, this report would not have been possible without the contributions of multiple team members. We thank Sunny Cao and Rebecca Goetz for their excellent data management and analytic support. We also appreciate the contributions of Charles Harding and Bonnee Groover to the editing and production of the report.

**Suggested citation:**
Princiotta, D., & Caspary, K. (2022). *YouthTruth concurrent and predictive validity study*. SRI International.

# Contents

# Tables

# Executive Summary

## Overview of the Study

This study tested the relationship at the school level between YouthTruth student survey scales and key academic and behavioral outcomes, providing clear evidence of the concurrent and predictive validity of these scales. Six YouthTruth student experience scales were tested:

- Academic challenge
- Culture
- Engagement
- Relationships with teachers
- Belonging and peer collaboration
- College and career readiness

**School-Level Outcomes**

The study examined five outcomes to validate the student experience scales.

***Academic outcomes:***

- Reading proficency rates
- Math proficiency rates
- Ninth-grade retention rates

***Behavioral outcomes:***

- Chronic absence rates
- Suspensions

Four student experience scales—academic challenge, culture, engagement, and relationships with teachers—are available for all three school levels: elementary, middle, and high school. The secondary survey includes two additional student experience scales—belonging and peer collaboration, and college and career readiness (the latter only asked of high school students). Based on the study results, school and district staff can be confident that school-level results on all YouthTruth student experience scales are associated with key school-level outcomes of interest.

## Background on the YouthTruth Student Survey

YouthTruth is a national survey project that harnesses student and stakeholder feedback to help guide decision-making by school leaders and education funders. The project began in 2008 as a collaboration between the Center for Effective Philanthropy and the Bill and Melinda Gates Foundation to provide a means to hear directly and systematically from students. YouthTruth is based on the premise that, whether you are a teacher, a principal, a superintendent, a nonprofit leader, or a funder, getting timely feedback from those you are trying to help, and listening to that feedback, enables improvement. To date, YouthTruth has surveyed over 2 million students across 39 states.

## Motivation for the Study

Listening to students makes sense because their feedback has been empirically linked to desired student outcomes—school climate measures based on student surveys have been shown to be leading indicators of student outcomes such as grades and graduation rates (see, for example,

Berkowitz et al., 2017). With more real-time feedback from students, school leaders can focus on improving aspects of school climate that they know to be associated with these desired outcomes.

This study establishes the concurrent and predictive validity of the YouthTruth student experience scales. Previous research has provided confirmatory evidence that the YouthTruth student experience scales are internally consistent and that the survey items making up those scales are well aligned with specified constructs (YouthTruth, 2018). The analyses for this study built on that evidence by examining whether the student experience scales were associated with important school-level academic and behavioral outcomes, including reading and math proficiency rates, ninth-grade retention rates, chronic absence rates, and suspensions.

## Research Questions

The study focused on two primary research questions related to the concurrent and predictive validity of the YouthTruth student experience scales.

**(1) Concurrent validity:**

1A. To what extent are 2018 YouthTruth student experience scales associated with 2018 reading proficiency rates, math proficiency rates, chronic absence rates, and suspensions at the school level, after controlling for school demographics and state?

1B. To what extent are 2018 ninth-grade YouthTruth student experience scales associated with ninth-grade retention rates at the end of the school year?

**(2) Predictive validity:**

To what extent are 2017 YouthTruth student experience scales associated with 2018 reading proficiency rates, math proficiency rates, chronic absence rates, and suspensions at the school level, after controlling for school demographic factors and state?

## Analytic Sample

The analytic samples for this study included public schools at the elementary, middle, and high school levels that participated in the YouthTruth student survey in 2018 (concurrent validity) or 2017 and 2018 (predictive validity). Of these public schools, 87% were traditional public schools and 13% were public charter schools.

The actual sample for each analysis varied based on availability of the outcome and data quality. The largest concurrent validity sample at each school level was for proficiency rates: 170 elementary schools, 110 middle schools, and 110 high schools (Table ES-1). The predictive validity sample excluded schools that did not participate in the YouthTruth student survey in both 2017 and 2018; thus, it was smaller than the concurrent validity sample for all school levels.

*Table ES-1. Range of analytic samples by analysis type and school level*

| Analysis type | Elementary schools | Middle schools | High schools |
|---|---|---|---|
| **Concurrent validity** | 150–170 | 100–110 | 30–110 |
| **Predictive validity** | 130–150 | 70–110 | 30–80 |

*Note.* Samples rounded to nearest 10.

*Source.* SRI Education analyses of data from the YouthTruth student survey and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

## Analytic Approach

The study team examined both concurrent and predictive validity, controlling for aggregate student characteristics at the school level to account for differences in student composition from one school to another. The analytic models also included fixed effects for state to account for differences in assessments and proficiency thresholds used from state to state. For concurrent validity analyses, the team examined outcomes measured in 2018 with predictors of interest and control variables measured in 2018. For predictive validity analyses, the team examined outcomes measured in 2018 on predictors of interest and control variables measured in 2017.

### Ninth-Grade Retention

Validity analyses involving ninth-grade retention rates used survey scales collected in 2018 for ninth-grade respondents only as predictors of ninth-grade retention at the completion of the 2018 school year. Because the survey measure and outcome were from the same year, the study team classified these ninth-grade retention analyses as concurrent validity. Examining predictive validity for ninth-grade retention (i.e., the relationship between the survey responses of 2018 ninth graders and the retention rate for the 2019 ninth-grade class) would not make sense.

## Overview of the Results

> Each YouthTruth student experience scale—academic challenge, culture, engagement, relationships with teachers, belonging and peer collaboration, and college and career readiness—demonstrated statistically significant evidence of concurrent validity for at least one outcome of interest at one or more school levels, after controlling for school demographics and state.
>
> Further, each YouthTruth student experience scale demonstrated statistically significant evidence of predictive validity for at least one outcome of interest at one or more school levels, after controlling for school demographics and state.

Table ES-2 summarizes statistically significant results in the expected direction only: positive for reading and math proficiency rates and negative for retention and chronic absence rates and suspensions. The expected relationship between the student experience scales and reading and math proficiency rates was positive: higher student experience scale scores were expected to be

associated with higher proficiency rates. Lower rates are more desirable for the other three outcomes, so the expected relationship was reversed: higher student experience scores were expected to be associated with lower ninth-grade retention rates, chronic absence rates, and suspension rates. A few broad trends stand out.

- **YouthTruth student experience scales more often demonstrated concurrent than predictive validity.** This pattern makes sense because alignment between students contributing to the student experience scales and the outcome measures is tighter (more likely to constitute the same students) within the same year.
- **The student experience scales more often demonstrated concurrent or predictive validity for the academic and behavioral outcomes investigated in middle and high school than in elementary school.** One possible reason for this is that fewer response options on elementary school survey items may lead to less variation in the scales.
- **Three student experience scales— academic challenge, culture, and belonging and peer collaboration—had relatively large numbers of statistically significant relationships with key academic and behavioral outcomes across school levels.** In contrast, two scales—engagement and relationships with teachers—demonstrated concurrent validity primarily for behavioral outcomes. The college and career readiness scale, which was only collected in high school, demonstrated concurrent validity only for suspensions.
- **The student experience scales were associated with improved behavioral outcomes.**
  - **All six scales were associated with lower suspensions.** The scales demonstrated concurrent validity with suspensions, often for multiple school levels—or even all school levels in the case of academic challenge—and the scales demonstrated predictive validity for one school level (typically high school).
  - **Almost all the scales were associated with lower chronic absence rates.** Five of the six scales—all except college and career readiness—were associated with lower chronic absence rates at one school level.
- **Many student experience scales were associated with reading proficiency rates, but few were associated with ninth-grade retention or math proficiency rates.**
  - **Most scales were associated with higher reading proficiency rates.** Four of the six scales—all except engagement and college and career readiness—were associated with increased reading proficiency rates.
  - **Only the culture scale was associated with lower ninth-grade retention rates.** The small number of schools in the ninth-grade retention analytic samples

means that this analysis is underpowered and may not detect additional underlying relationships.

- **Few scales were associated with increased math proficiency rates.** Only the culture scale and the belonging and peer collaboration scale demonstrated evidence of concurrent validity with math proficiency rates.

***Table ES-2. Overview of concurrent (✓) and predictive (+) validity of YouthTruth student experience scales by academic and behavioral outcomes and school level***

| | | Academic outcomes | | | Behavioral outcomes | |
|---|---|---|---|---|---|---|
| Student experience scale | School level | Reading proficiency rate | Math proficiency rate | Ninth-grade retention rate[a] | Chronic absence rate | Log suspension index |
| Academic challenge | Elementary | | | NA | | ✓ |
| | Middle | ✓ + | | NA | ✓ | ✓ |
| | High | ✓ + | | | | ✓ + |
| Culture | Elementary | ✓ | ✓ | NA | | |
| | Middle | ✓ + | ✓ | NA | ✓ | ✓ |
| | High | | | ✓ | | ✓ + |
| Engagement | Elementary | | + | NA | ✓ | |
| | Middle | | | NA | | ✓ |
| | High | | | | | ✓ + |
| Relationships with teachers | Elementary | | | NA | | |
| | Middle | ✓ + | | NA | ✓ | ✓ |
| | High | | | | | ✓ + |
| Belonging and peer collaboration[b] | Middle | ✓ + | | NA | | ✓ + |
| | High | ✓ + | ✓ | | ✓ + | |
| College and career readiness[c] | High | | | | | ✓ + |

✓ = statistically significant ($p < 0.05$) evidence of concurrent validity with relationship in expected direction.

\+ = statistically significant ($p < 0.05$) evidence of predictive validity with relationship in expected direction.

NA = not applicable.

*Note.* Blank cells represent no statistically significant relationship detected in the expected direction, after controlling for school demographics and state.

[a] Examined concurrently validity only for ninth-grade retention.

[b] Belonging and peer collaboration scale not included on elementary school survey.

[c] College and career readiness scale only included on high school survey.

*Source.* SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use EDFacts data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

The following paragraphs summarize the observed relationships for each YouthTruth student experience scale, highlighting the outcomes with associations at multiple school levels or for both concurrent and predictive validity.

***Academic Challenge.*** The academic challenge scale demonstrated both concurrent and predictive validity with eighth-grade and high school reading proficiency rates and concurrent validity with suspensions at all school levels (plus predictive validity for high school). In addition, the scale demonstrated concurrent validity for middle school chronic absence rates.

***Culture.*** The culture scale demonstrated concurrent validity with fifth- and eighth-grade reading and math proficiency rates (plus predictive validity with eighth-grade reading proficiency rates) and with middle and high school suspensions (plus predictive validity with high school suspensions). In addition, the scale demonstrated concurrent validity with ninth-grade retention rates and the middle school chronic absence rates.

***Engagement.*** The engagement scale demonstrated concurrent validity with the elementary school chronic absence rates and with the middle and high school suspensions (plus predictive validity with high school suspensions). In addition, the scale demonstrated predictive validity for fifth-grade math proficiency rates (but not concurrent validity) and showed an unexpected negative predictive relationship with middle school math proficiency rates (not shown in Table 5). Given the contradictory nature of these two findings and the lack of concurrent validity findings related to middle school math proficiency rates, this negative relationship may be anomalous.

***Relationships with Teachers.*** The relationships with teachers scale demonstrated concurrent and predictive validity with eighth-grade reading proficiency rates. The scale also demonstrated concurrent validity for suspensions in middle and high school (as well as predictive validity in high school) and for middle school chronic absence rates.

***Belonging and Peer Collaboration.*** The belonging and peer collaboration scale demonstrated concurrent and predictive validity with eighth-grade and high school reading proficiency rates and with high school chronic absence rates and middle school suspension rates. In addition, it demonstrated concurrent validity with high school math proficiency rates (the only student experience scale to do so).

***College and Career Readiness.*** The college and career readiness scale, only available for high school students, demonstrated concurrent and predictive validity for suspensions. The scale also demonstrated concurrent validity in an unexpected direction with high school reading proficiency rates: Higher college and career readiness scores were associated with lower reading proficiency rates. This finding is counterintuitive, as higher skills in English language arts are typically associated with increased college and career readiness among students. One possible explanation (among many) is related to the scale's focus on career over college readiness. High schools with low reading proficiency rates may tend to emphasize career over college preparation.

# Conclusion

This study has made a substantial contribution to the literature by showing that each YouthTruth student experience scale demonstrates concurrent validity for one or more key academic and behavioral outcomes in at least one school level. The same holds for predictive validity. These findings support the use of the YouthTruth student experience scales in school improvement efforts.

Several factors may have limited the study team's ability to fully detect associations between the scales and school-level outcomes, including small sample sizes, the use of school-level rather than student-level data, and differences in the composition of students contributing to the student experience scales and the outcomes for each school. Together, these factors mean that additional associations between YouthTruth student experience scales and key outcomes may exist, beyond those reported for this study. Future research could examine additional student or school-level outcomes of interest and use student-level data to generate more precise estimates of the relationships between the student experience scales and these outcomes.

# Chapter I. Introduction

## Overview of the Study

The study objective was to test the concurrent and predictive validity of the YouthTruth student experience scales measuring academic challenge, culture, engagement, relationships with teachers, belonging and peer collaboration, and college and career readiness.[1] The study specifically examined the relationship at the school level between the student experience scales and key academic and behavioral outcomes.

## Background on the YouthTruth Student Survey

YouthTruth is a national survey project that harnesses student and stakeholder feedback to help school leaders and education funders make better decisions that lead to better outcomes for students. The project began in 2008 as a collaboration between the Center for Effective Philanthropy and the Bill and Melinda Gates Foundation to provide a means to hear directly and systematically from students. Members of the founding organizations believed that the voices of students, who are at the very center of our education system, were too often missing from the conversation about improving education outcomes and experiences.

YouthTruth sought to respond to this problem by bringing rigorously collected and comparatively presented student perceptions to education funders as well as to the leaders of their partner schools and districts. YouthTruth is based on the premise that, whether you are a teacher, a principal, a superintendent, a nonprofit leader, or a funder, getting timely feedback from those you are trying to help, and listening to that feedback, enables improvement. To date, YouthTruth has surveyed over 2 million students across 39 states.

## Motivation for the Study

YouthTruth believes that the people at the center of schools—the students—should have a voice in how their school works for them and how it can work better. Further, listening to students makes sense because student feedback has been empirically linked to desired student outcomes—school climate measures based on student surveys have been shown to be leading indicators of student outcomes such as grades and graduation rates (see, for example, Berkowitz et al., 2017). With more real-time feedback from students, school leaders can focus on improving aspects of school climate that they know to be associated with these desired outcomes.

[1] A supplemental objective of the study was to investigate the concurrent and predictive validity of individual survey items about academic supports, extracurricular activities, student expectations, bullying, and out-of-school obstacles to student success. This objective was addressed through supplemental analyses, the results of which are available in the online technical supplement.

This study built on earlier studies of YouthTruth student survey scales. Previous research has provided confirmatory evidence that the YouthTruth student experience scales are internally consistent and that the survey items making up those scales are well aligned with specified constructs (YouthTruth, 2018). To further validate the student experience scales, this study examined their concurrent and predictive validity. These analyses provided evidence on whether the student experience scales were related with important school-level academic and behavioral outcomes, including reading and math proficiency rates, ninth-grade retention rates, chronic absence rates, and suspensions. If school-level results on the YouthTruth student survey are associated with key outcomes of interest, school and district staff can more confidently use the YouthTruth student experience scales to inform their school improvement efforts.

## Research Questions

The study focused on two primary research questions related to the concurrent and predictive validity of the YouthTruth student experience scales.

**(1) Concurrent validity:**

1A. To what extent are 2018 YouthTruth student experience scales associated with 2018 reading proficiency rates, math proficiency rates, chronic absence rates, and suspensions at the school level, after controlling for school demographics and state?

1B. To what extent are 2018 ninth-grade YouthTruth student experience scales associated with ninth-grade retention rates at the end of the school year?

**(2) Predictive validity:**

To what extent are 2017 YouthTruth student experience scales associated with 2018 reading proficiency rates, math proficiency rates, chronic absence rates, and suspensions at the school level, after controlling for school demographic factors and state?

The study team examined both concurrent and predictive validity because the approaches have different advantages and limitations. The analytic samples for concurrent validity were larger than those for predictive validity, which increased the team's ability to detect statistically significant findings (see the Analytic Approach section for more details). Further, because the survey scale and outcomes were measured in the same year for the concurrent validity analyses, the composition of students in each school that contributed to the student experience scales was better aligned with the composition of students that made up the school-level outcomes estimates for concurrent than predictive validity. Nonetheless, the study examined predictive validity in addition to concurrent validity because students are typically exposed to school climate over multiple years, and school climate in one year could plausibly affect student outcomes in the subsequent year.[2]

[2] A third exploratory research question focused on changes over time: To what extent are changes from 2017 to 2018 in YouthTruth student experience scales associated with changes in school outcomes from 2017 to 2018, after controlling for

# Chapter II. Overview of Data and Methods

## Data Sources

The analytic data set for the study combined data aggregated to the school level from three sources:

- YouthTruth student survey data from all participating schools in 2017 and 2018.
- U.S. Department of Education restricted-use ED*Facts* data from 2018.
- Office of Civil Rights 2018 Civil Rights Data Collection (CRDC).

To construct the analytic data set, the study team started with a data file containing 2017 and 2018 survey results aggregated to the school level from all schools participating in the YouthTruth student survey in those years. The team appended this file with 2018 restricted-use ED*Facts* data, including student demographics and academic outcome data aggregated to the school level. Finally, the ninth-grade retention data and behavioral outcomes for the study were from the 2018 CRDC.

### Outcomes of Interest

The study team examined five aggregate school-level outcomes (Table 1). Two of the first three—state assessment results in English language arts and math—were from the restricted-use ED*Facts* data file. The team used fifth-grade, eighth-grade, and high school proficiency rates to validate each corresponding school-level survey (elementary, middle, and high school). Ninth-grade retention and the two behavioral outcomes—chronic absences and suspensions—came from the CRDC.[3] For these three outcomes, lower values are desirable. For example, a low ninth-grade retention rate means that a low proportion of ninth graders were required to repeat ninth grade.

**Suspensions**

For the study, suspensions was defined as the natural log of the suspension index (Table 1). The suspension index is similar to the suspension rate (the percentage of students in a school receiving suspensions in a given year), but students with both in- and out-of-school suspensions are double counted.

Because schools typically had low suspension indices but some had very large ones, the study team calculated the natural log of the suspension index before running statistical models to create a suspensions variable with a more normal distribution.

changes in school demographic factors? Due to extremely large standard errors associated with small sample sizes for the change-over-time analyses, results from these analyses are not presented or discussed in the main body of this report. However, those results are available in the online technical supplement.

[3] The study team also examined an additional academic outcome from the ED*Facts* data file: the adjusted cohort high school graduation rate. Results for this analysis are available in the online technical supplement but are not highlighted in this report because of the greater alignment of ninth-grade retention in determining the relationship between YouthTruth indicators and progression in school.

*Table 1. Outcomes definitions and data sources*

| Outcome | Description | Data source |
|---|---|---|
| Academic outcomes | | |
| Reading proficiency rate | Percentage of fifth graders (elementary), eighth graders (middle) and high school students (grades vary) who score proficient on the state English language arts test in a given year. | EDFacts |
| Math proficiency rate | Percentage of fifth graders (elementary), eighth graders (middle) and high school students (grades vary) who score proficient on the state math test in a given year. | EDFacts |
| Ninth-grade retention rate | Percentage of ninth graders who are retained in ninth grade at the end of the year (i.e., held back in ninth grade because they are not ready to advance to tenth grade). | CRDC |
| Behavioral outcomes | | |
| Chronic absence rate | Percentage of students absent for 10% or more of the school days in a given year. | CRDC |
| Suspensions | Natural log of the suspension index, the sum of the number of students receiving in-school suspensions and the number of students receiving out-of-school suspensions divided by the total student enrollment. | CRDC |

CRDC = Civil Rights Data Collection.

## Predictors of Interest

The YouthTruth student survey collects information on student perceptions to characterize essential elements of school culture and climate in both elementary and secondary schools. These elements are measured with scales based on student responses to specific YouthTruth survey questions with Likert scale response options. Through psychometric testing, including confirmatory factor analysis, these student experience scales have been shown to be internally consistent, valid, and reliable (YouthTruth, 2018).

This study investigated the concurrent and predictive validity of six student experience scales (Table 2).[4] The student survey instruments, and thus the student experience scales included, vary by school level (elementary vs. secondary). For example, the culture scale focuses on classroom culture for elementary and school culture for secondary. Further, the Likert scale on the elementary school survey has fewer response options (3) than on the middle and high school survey (5). Four student experience scales are available for all three school levels—academic challenge, culture, engagement, and relationships with teachers—but the definition and items that constitute each scale may differ for the elementary and secondary versions. Items on the elementary version tend to be framed in terms of students' perceptions of their teachers, whereas more items on the secondary version directly ask students about themselves and their

[4] The YouthTruth student survey includes additional items that collect information on key student perceptions, including participation in academic supports, extracurricular activities, and nonschool barriers to learning. Further, the elementary student survey includes an instructional methods scale. The study team examined these individual items and the instructional methods scale to determine concurrent and predictive validity. The results of these analyses are available in the online technical supplement.

perceptions of academic challenge, engagement, and relationships with teachers. The secondary version includes two additional student experience scales: belonging and peer collaboration; and college and career readiness (only asked of high school students). Based on prior research, Cronbach's alphas for the student experience scales ranged between 0.74 (belonging and peer collaboration) and 0.88 (college and career readiness) for high school student experience scales and between 0.66 (engagement) and 0.83 (academic challenge) for middle school student experience scales (YouthTruth, 2018).

### *Table 2. Descriptions of the YouthTruth student experience scales*

| Scale | Description (grade level) |
|---|---|
| Academic challenge | **Academic expectations (elementary)**<br>The teacher poses challenging and substantive work to students, building a strong academic work ethic and critical thinking skills. |
| | **Academic challenge (middle and high school)**<br>Students feel they are challenged by their coursework and teachers. |
| Culture | **Classroom culture (elementary)**<br>The teacher develops a classroom environment premised on respect, motivation, and organization. |
| | **School culture (middle and high school)**<br>Students believe that their school fosters a culture of respect and fairness. |
| Engagement | **Student engagement**<br>Students perceive themselves as engaged with their school and their education. |
| Relationships with teachers | **Personal relationships (elementary)**<br>The teacher supports students' academic success through positive interpersonal interactions. |
| | **Teacher support (middle and high school)**<br>Students feel they receive support and personal attention from their teachers. |
| Belonging and peer collaboration | **Relationships with peers (middle and high school)**<br>Students have supportive, collaborative relationships with their classmates. |
| College and career readiness | **Postsecondary planning (high school only)**<br>Students feel their school equipped them to pursue college and careers. |

*Source*. YouthTruth (2018).

Each student experience scale score was created by calculating an average score across the items in each measure for each respondent. The study team then aggregated these composite scores to the school level by calculating the mean score across respondents from all grade levels in the school (e.g., third through fifth, sixth through eighth, and ninth through twelfth).[5] The team also aggregated the survey responses of ninth graders in each high school for the ninth-grade retention analyses. Each student experience scale was standardized within school level for analysis.

[5] The elementary school survey is first administered in third grade.

### Control Variables

In addition to the predictors of interest, the study team included state as well as the following school demographic variables in the analyses:

- Percent economically disadvantaged
- Percent English learner
- Percent with a disability
- Percent Asian, non-Hispanic
- Percent Black, non-Hispanic
- Percent Hispanic
- Percent other race, non-White and non-Hispanic
- Log school enrollment
- Percent male

Additional details on the construction of variables used in the analyses are included in the online technical supplement.

## Analytic Sample

The analytic sample for this study included public schools at the elementary, middle, and high school level that participated in the YouthTruth student survey in 2017, 2018, or both. Of these public schools, 87% were traditional public schools and 13% were public charter schools. The study team excluded private schools and fully virtual schools, as well as any schools missing demographic data from ED*Facts*.

### Sample Variation by Analysis

The actual sample for each analysis varied based on availability of the outcome and data quality. The largest concurrent validity sample at each school level was for proficiency rates: 170 elementary schools, 110 middle schools, and 110 high schools (Table 3).[6] The predictive validity sample included schools that participated in YouthTruth in 2017 and was smaller than the concurrent validity sample for all school levels. For chronic absence rate analyses, it was unclear whether zeros in the CRDC data file were true zeros or whether they represented missing data. Further, more than 10 schools in a single state reported unusually high chronic absence rates—above 50%. As a result, the study team excluded schools with chronic absence rates of 0% or above 50% from the chronic absence rate analyses. Similarly, the team excluded schools that reported zero suspensions from the suspensions analyses and schools that reported zero ninth-grade retentions from the ninth-grade retention rate analyses. After these exclusions, the concurrent validity sample dropped to 70 schools for middle school chronic absence rate

[6] All sample sizes in this report and in the online technical supplement are rounded to the nearest 10, following the U.S. Department of Education's restricted-use data regulations.

analyses, to 70 schools for high school chronic absence rate analyses, and to 30 schools for ninth-grade retention rate analyses.

***Table 3. Analytic samples by analysis type and school level***

| Analysis type | School level | Reading proficiency | Math proficiency | Ninth-grade retention | Chronic absence | Suspensions |
|---|---|---|---|---|---|---|
| Concurrent validity | Elementary | 170 | 170 | NA | 170 | 150 |
| | Middle | 110 | 110 | NA | 100 | 100 |
| | High | 110 | 110 | 30 | 80 | 100 |
| Predictive validity | Elementary | 150 | 150 | NA | 110 | 130 |
| | Middle | 110 | 110 | NA | 70 | 100 |
| | High | 80 | 70 | NA | 70 | 80 |

NA = not applicable.

*Note*. Sample sizes are rounded to the nearest 10.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016-2018).

The limited sample sizes have implications for the generalizability of the findings and for statistical power—the ability to detect true underlying relationships between the YouthTruth student experience scales and the outcomes of interest. The full analytic sample is not generalizable to all U.S. public schools because schools that chose to participate in the YouthTruth student survey were not a random selection of schools. The reduction in samples for specific outcomes further limited the generalizability of the findings; the results do not generalize to those schools that truly have no suspensions, no chronic absences, and no students retained in ninth grade. In addition, some analyses were low in statistical power. For example, ninth-grade retention analyses were based on just 30 schools. As a result, the study team's estimates of the relationships between the YouthTruth student experience scales and the outcomes of interest were relatively imprecise, with large confidence intervals, increasing the probability that the study failed to detect underlying associations. The risk of missing a true underlying association is greater for analyses with smaller sample sizes. In this study, the sample size for predictive validity analyses based on the same outcomes was smaller for concurrent validity analyses and the sample size for middle and high school analyses was smaller than for elementary school.

## Sample Descriptives

Table 4 shows descriptive statistics for the elementary, middle, and high school analytic samples for the largest samples—the reading proficiency rate samples for concurrent validity analyses. These samples included schools in 12 states for the elementary school analysis, 15 states for the middle school analysis, and 22 states for the high school analysis. The largest proportion of schools (44% in the elementary school analysis, 43% in the middle school analysis, and 47% in

the high school analysis) were located in California. Further, 25% of elementary schools and around 20% of middle and high schools in the samples were located in Texas. The school-level average percentage of English language arts test-takers who identified as Hispanic was more than 35% at all school levels (39% in fifth grade, 41% in eighth grade, and 42% in high school). Although in 2018, 27% of public school students nationwide identified as Hispanic, over 50% identified as Hispanic in California and Texas (National Center for Education Statistics, 2020). In addition, the school-level average percentage of English language arts test-takers who were economically disadvantaged was 60% in elementary and middle school and 55% in high school. Descriptives specific to other analyses are available in the online technical supplement.

### *Table 4. Descriptive Statistics for the Concurrent Validity Reading Proficiency Rate Samples*

| | Elementary | | | Middle | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| Reading proficiency rate | 170 | 53.62 | 18.15 | 110 | 49.24 | 17.88 | 110 | 52.50 | 23.66 |
| Number of students tested | 170 | 77.20 | 28.97 | 110 | 217.24 | 144.72 | 110 | 290.75 | 300.28 |
| Log number of students tested | 170 | 4.26 | 0.44 | 110 | 5.07 | 0.90 | 110 | 5.16 | 1.10 |
| Percent with a disability | 170 | 14.22 | 7.23 | 110 | 12.71 | 6.18 | 110 | 11.52 | 5.49 |
| Percent economically disadvantaged | 170 | 60.35 | 28.25 | 110 | 59.48 | 27.97 | 110 | 55.14 | 25.50 |
| Percent English learner | 170 | 26.51 | 23.96 | 110 | 17.05 | 16.95 | 110 | 14.48 | 14.31 |
| Percent homeless | 170 | 2.60 | 4.23 | 110 | 2.66 | 2.97 | 110 | 4.28 | 7.00 |
| Percent Asian, non-Hispanic | 170 | 13.52 | 19.55 | 110 | 8.89 | 14.73 | 110 | 5.89 | 8.31 |
| Percent black, non-Hispanic | 170 | 14.38 | 23.22 | 110 | 12.46 | 22.43 | 110 | 10.53 | 17.50 |
| Percent Hispanic | 170 | 38.57 | 27.99 | 110 | 41.29 | 30.29 | 110 | 42.41 | 29.75 |
| Percent male | 170 | 51.49 | 6.12 | 110 | 50.44 | 9.39 | 110 | 51.76 | 7.28 |
| California | 170 | 0.44 | 0.50 | 110 | 0.43 | 0.50 | 110 | 0.47 | 0.50 |
| District of Columbia | 170 | 0.03 | 0.18 | 110 | 0.04 | 0.21 | 110 | 0.02 | 0.14 |
| Florida | | | | | | | 110 | 0.01 | 0.10 |
| Hawaii | | | | 110 | 0.01 | 0.09 | 110 | 0.01 | 0.10 |
| Iowa | 170 | 0.01 | 0.08 | | | | | | |
| Kansas | 170 | 0.01 | 0.08 | 110 | 0.01 | 0.09 | 110 | 0.01 | 0.10 |
| Louisiana | | | | | | | 110 | 0.00 | 0.00 |
| Massachusetts | | | | | | | 110 | 0.02 | 0.14 |
| Maine | | | | | | | 110 | 0.01 | 0.10 |
| Michigan | 170 | 0.00 | 0.00 | 110 | 0.00 | 0.00 | 110 | 0.00 | 0.00 |
| Minnesota | 170 | 0.01 | 0.08 | 110 | 0.01 | 0.09 | 110 | 0.01 | 0.10 |
| North Carolina | | | | 110 | 0.00 | 0.00 | 110 | 0.01 | 0.10 |
| New Jersey | | | | 110 | 0.01 | 0.09 | 110 | 0.01 | 0.10 |
| New York | | | | 110 | 0.01 | 0.09 | 110 | 0.02 | 0.14 |
| Ohio | 170 | 0.07 | 0.25 | 110 | 0.06 | 0.24 | 110 | 0.08 | 0.27 |
| Oklahoma | | | | | | | 110 | 0.00 | 0.00 |
| Oregon | 170 | 0.01 | 0.08 | 110 | 0.10 | 0.30 | 110 | 0.05 | 0.21 |
| Pennsylvania | 170 | 0.06 | 0.24 | 110 | 0.03 | 0.16 | 110 | 0.03 | 0.17 |
| Tennessee | | | | | | | 110 | 0.00 | 0.00 |
| Texas | 170 | 0.25 | 0.44 | 110 | 0.21 | 0.41 | 110 | 0.20 | 0.40 |
| Washington | 170 | 0.11 | 0.31 | 110 | 0.08 | 0.27 | 110 | 0.05 | 0.21 |
| Wisconsin | 170 | 0.01 | 0.08 | 110 | 0.01 | 0.09 | 110 | 0.01 | 0.10 |

# Analytic Approach

The study team examined the relationship between the YouthTruth student experience scales (four elementary, five middle, and six high school; see Table 2) and both academic and behavioral outcomes. The team assessed concurrent and predictive validity for each scale separately in elementary, middle, and high school. The analytic models controlled for aggregate student characteristics at the school level to account for differences in student composition from one school to another, as well as fixed effects for state to account for differences in assessments and proficiency thresholds used from state to state. For concurrent validity analyses, the team fit ordinary least squares (OLS) regressions of outcomes measured in 2018 on control variables and standardized predictors of interest measured in 2018. For predictive validity analyses, the team fit OLS regressions of outcomes measured in 2018 on control variables and standardized predictors of interest measured in 2017. For more details on the analytic models, see the online technical supplement.

### Ninth Grade Retention

Our validity analyses involving ninth-grade grade retention rates used survey scales collected in 2018 for ninth grade respondents only as predictors of 9th grade retention at the completion of the 2018 school year. Because the survey measure and outcome are from the same year, we classify these ninth-grade retention analyses as concurrent validity; examining predictive validity for ninth grade retention (i.e., the relationship between the survey responses of 2018 ninth graders and the retention rate for the 2019 ninth-grade class) would not make sense.

Three aspects of the analytic approach limited the study team's ability to detect underlying relationships between the YouthTruth student experience scales and outcomes: sample size, alignment of scale and outcome student composition, and reliance on school-level data. The sample size limitation was most acute for the ninth-grade retention analysis (see the Sample Variation by Analysis section). The alignment between the students contributing to the student experience scales and the outcomes in each school varied by analysis. Because the composition of students in a school changes from year to year, as cohorts turn over and individual students move to other schools, students comprising the survey scale and outcomes in each school are more likely to be the same within the same school year than in adjacent years, i.e., for concurrent validity analyses than for predictive validity analyses. Further, this alignment was greater for ninth-grade retention rates, chronic absence rates, and the suspensions analyses than it was for reading and math proficiency rates analyses. For example, in middle schools, the study team examined the relationship between student experience scales for all middle school grade levels with eighth-grade reading or math proficiency scores, and took a parallel approach for elementary (fifth-grade reading or math proficiency rates) and high school. In contrast, for chronic absence rates and suspensions, the team examined the relationship between the student experiences scales for all students with the outcomes for all students in each school. On the basis

of the degree of scale and outcome alignment, the team expected to see a greater number of statistically significant results for models with more tightly aligned samples—that is, for concurrent validity than for predictive validity, and for ninth-grade retention and the two behavioral outcomes than for proficiency rates. Finally, the reliance of analyses on school-level rather than student-level data decreased the precision of estimates, limiting the ability to detect relationships across analyses. Despite these limitations, the study team's analyses found multiple statistically significant relationships between student experience scales and desired outcomes.

# Chapter III. Results

This chapter begins with an overview of high-level findings and follows with a more detailed presentation of findings related to each YouthTruth student experience scale. The overview indicates whether any statistically significant evidence of concurrent or predictive validity in the expected direction was found. The second part presents detailed findings on the magnitude of the relationships between the student experience scales and academic and behavioral outcomes of interest by scale, including statistical significance.

## Overview of the Results

> Each YouthTruth student experience scale—academic challenge, culture, engagement, relationships with teachers, belonging and peer collaboration, and college and career readiness—demonstrated statistically significant evidence of concurrent validity for at least one outcome of interest at one or more school levels, after controlling for school demographics and state.
>
> Further, each YouthTruth student experience scale demonstrated statistically significant evidence of predictive validity for at least one outcome of interest at one or more school levels, after controlling for school demographics and state.

Table 5 summarizes statistically significant results *in the expected direction only*: positive for reading and math proficiency rates and negative for retention, chronic absence rates and suspensions.[7] The expected relationship between the student experience scales and reading and math proficiency rates was positive: higher student experience scale scores were expected to be associated with higher proficiency rates. Lower rates are more desirable for the other three outcomes, so the expected relationship was reversed: higher student experience scores were

[7] The study team observed two statistically significant relationships between the YouthTruth student experience scales and outcomes in an unexpected direction. College and career readiness was associated with lower same-year high school reading proficiency rates, and engagement was associated with lower following-year eighth-grade math proficiency rates. These findings did not form any consistent patterns across grade levels or concurrent and predictive validity.

expected to be associated with lower ninth-grade retention rates, chronic absence rates, and suspension rates. A few broad trends stand out.

- **YouthTruth student experience scales more often demonstrated concurrent than predictive validity.** This pattern makes sense because alignment between students contributing to the student experience scales and the outcome measures is tighter (more likely to constitute the same students) within the same year.
- **The student experience scales more often demonstrated concurrent or predictive validity for the academic and behavioral outcomes investigated in middle and high school than in elementary school.** One possible reason for this is that fewer response options on elementary school survey items may lead to less variation in scales.
- **Three student experience scales— academic challenge, culture, and belonging and peer collaboration—had relatively large numbers of statistically significant relationships with key academic and behavioral outcomes across school levels.** In contrast, two scales—engagement and relationships with teachers—demonstrated concurrent validity primarily for behavioral outcomes. The college and career readiness scale, which was only collected in high school, demonstrated concurrent validity only for suspensions.
- **The student experience scales were associated with improved behavioral outcomes.**
  - **All six scales were associated with lower suspensions.** The scales demonstrated concurrent validity with suspensions, often for multiple school levels—or even all school levels in the case of academic challenge—and the scales demonstrated predictive validity for one school level (typically high school).
  - **Almost all the scales were associated with lower chronic absence rates.** Five of the six scales—all except college and career readiness—were associated with lower chronic absence rates at one school level.
- **Many student experience scales were associated with reading proficiency rates, but few were associated with ninth-grade retention or math proficiency rates.**
  - **Most scales were associated with higher reading proficiency rates.** Four of the six scales—all except engagement and college and career readiness—were associated with increased reading proficiency rates.
  - **Only the culture scale was associated with lower ninth-grade retention rates.** The small number of schools in the ninth-grade retention analytic samples means that this analysis is underpowered and may not detect additional underlying relationships.

- **Few scales were associated with increased math proficiency rates.** Only the culture scale and the belonging and peer collaboration scale demonstrated evidence of concurrent validity with math proficiency rates.

***Table 5. Overview of concurrent (✓) and predictive (+) validity of YouthTruth student experience scales by academic and behavioral outcomes and school level***

| Student experience scale | School level | Academic outcomes | | | Behavioral outcomes | |
|---|---|---|---|---|---|---|
| | | Reading proficiency rate | Math proficiency rate | Ninth-grade retention rate[a] | Chronic absence rate | Log suspension index |
| Academic challenge | Elementary | | | NA | | ✓ |
| | Middle | ✓ + | | NA | ✓ | ✓ |
| | High | ✓ + | | | | ✓ + |
| Culture | Elementary | ✓ | ✓ | NA | | |
| | Middle | ✓ + | ✓ | NA | ✓ | ✓ |
| | High | | | ✓ | | ✓ + |
| Engagement | Elementary | | + | NA | ✓ | |
| | Middle | | | NA | | ✓ |
| | High | | | | | ✓ + |
| Relationships with teachers | Elementary | | | NA | | |
| | Middle | ✓ + | | NA | ✓ | ✓ |
| | High | | | | | ✓ + |
| Belonging and peer collaboration[b] | Middle | ✓ + | | NA | | ✓ + |
| | High | ✓ + | ✓ | | ✓ + | |
| College and career readiness[c] | High | | | | | ✓ + |

✓ = statistically significant ($p < 0.05$) evidence of concurrent validity, after controlling for school demographics and state, with relationship in expected direction.

\+ = statistically significant ($p < 0.05$) evidence of predictive validity, after controlling for school demographics and state, with relationship in expected direction.

NA = not applicable.

*Note.* Blank cells represent no statistically significant relationship detected in the expected direction, after controlling for school demographics and state. Detailed information on variables and models are available in the Overview of Data and Methods section of the main report.

[a] Examined concurrently validity only for ninth-grade retention.

[b] Belonging and peer collaboration scale not included on elementary school survey.

[c] College and career readiness scale only included on high school survey.

*Source.* SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use EDFacts data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

The following paragraphs summarize the observed relationships for each YouthTruth student experience scale, highlighting the outcomes with associations at multiple school levels or for both concurrent and predictive validity.

***Academic Challenge.*** The academic challenge scale demonstrated both concurrent and predictive validity with eighth-grade and high school reading proficiency rates and concurrent validity with suspensions at all school levels (plus predictive validity for high school). In addition, the scale demonstrated concurrent validity for middle school chronic absence rates.

***Culture.*** The culture scale demonstrated concurrent validity with fifth- and eighth-grade reading and math proficiency rates (plus predictive validity with eighth-grade reading proficiency rates) and with middle and high school suspensions (plus predictive validity with high school suspensions). In addition, the scale demonstrated concurrent validity with ninth-grade retention rates and the middle school chronic absence rates.

***Engagement.*** The engagement scale demonstrated concurrent validity with the elementary school chronic absence rates and with the middle and high school suspensions (plus predictive validity with high school suspensions). In addition, the scale demonstrated predictive validity for fifth-grade math proficiency rates (but not concurrent validity) and showed an unexpected negative predictive relationship with middle school math proficiency rates (not shown in Table 5). Given the contradictory nature of these two findings and the lack of concurrent validity findings related to middle school math proficiency rates, this negative relationship may be anomalous.

***Relationships with Teachers.*** The relationships with teachers scale demonstrated concurrent and predictive validity with eighth-grade reading proficiency rates. The scale also demonstrated concurrent validity for suspensions in middle and high school (as well as predictive validity in high school) and for middle school chronic absence rates.

***Belonging and Peer Collaboration.*** The belonging and peer collaboration scale demonstrated concurrent and predictive validity with eighth-grade and high school reading proficiency rates and with high school chronic absence rates and middle school suspension rates. In addition, it demonstrated concurrent validity with high school math proficiency rates (the only student experience scale to do so).

***College and Career Readiness.*** The college and career readiness scale, only available for high school students, demonstrated concurrent and predictive validity for suspensions. The scale also demonstrated concurrent validity in an unexpected direction with high school reading proficiency rates: Higher college and career readiness scores were associated with lower reading proficiency rates. This finding is counterintuitive, as higher skills in English language arts are typically associated with increased college and career readiness among students. One possible explanation (among many) is related to the scale's focus on career over college readiness. High schools with low reading proficiency rates may tend to emphasize career over college preparation.

# Results by Survey Scale

In this more detailed presentation of results, we lead with concurrent validity, because the alignment between the YouthTruth student experience scales and outcomes samples is tighter than for predictive validity. The subsections for the six scales highlight relationships that are statistically significant for outcomes across multiple school levels and indicate when those relationships are present for predictive validity as well as concurrent validity.

## Academic Challenge

> With respect to academic outcomes, the YouthTruth academic challenge scale demonstrated statistically significant evidence of concurrent and predictive validity for eighth-grade and high school reading proficiency rates.
>
> Regarding behavioral outcomes, the academic challenge scale demonstrated statistically significant concurrent validity with respect to middle school chronic absence rates as well as suspensions at all school levels. In addition, the scale showed statistically significant predictive validity for suspensions in high school.

***The academic challenge scale demonstrated evidence of both concurrent and predictive validity with respect to eighth-grade and high school reading proficiency rates.***

In 2018, higher average scores on the academic challenge scale were associated with higher eighth-grade and high school reading proficiency rates, after controlling for school demographics and state (Table 6). Scoring 1 standard deviation higher on the academic challenge scale in 2018 was associated with a 3.7 percentage point increase in the 2018 eighth-grade reading proficiency rate and a 2.2 percentage point increase in the 2018 high school proficiency rate, on average.

Further, higher average scores on academic challenge scale in 2017 were associated with higher eighth-grade and high school reading proficiency rates in 2018, after controlling for school demographics and state. Scoring 1 standard deviation higher on the academic challenge scale in 2017 was associated with a 2.3 percentage point increase in the eighth-grade reading proficiency rate and a 4.8 percentage point increase in the high school reading proficiency rate in 2018, on average.

### *Table 6. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes of interest on the YouthTruth academic challenge scale by analysis type and school level*

| Outcome | Analysis type | Elementary | Middle | High |
|---|---|---|---|---|
| **Academic outcomes** | | | | |
| Reading proficiency rate | Concurrent | 1.54 | 3.67 ** | 2.17 * |
| | Predictive | 0.88 | 2.18 * | 4.81 * |
| Math proficiency rate | Concurrent | 1.32 | 3.15 | 2.05 |
| | Predictive | 1.30 | 0.19 | 1.54 |
| Ninth-grade retention rate | Concurrent | NA | NA | −0.27 |
| | Predictive | NA | NA | NA |
| **Behavioral outcomes** | | | | |
| Chronic absence rate | Concurrent | −0.17 | -2.11 ** | -0.45 |
| | Predictive | −0.54 | 2.74 | 1.79 |
| Log suspension index | Concurrent | −0.28 * | -0.37 ** | -0.33 ** |
| | Predictive | −0. | -0.14 | -0.45 ** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

*Note.* Detailed information on variables and models is available in the Overview of Data and Methods section.

*Source.* SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***Schools with higher levels of academic challenge had lower suspensions across school levels.*** In elementary, middle, and high school, higher levels of academic challenge were associated with lower same-year suspensions. Specifically, among schools with any suspensions in 2018, scoring 1 standard deviation higher on the academic challenge scale was associated with 24% lower suspension in elementary schools, 31% lower suspensions in middle schools, and 28% lower suspensions in high schools, on average.

Additionally, in high school, higher academic challenge levels in 2017 were associated with lower 2018 suspensions. Among high schools with any suspensions, a 1 standard deviation increase in school-level the academic challenge scale in 2017 was associated with an expected 36% decrease in the 2018 suspensions, on average.

Finally, higher levels of academic challenge were associated with lower rates of same-year chronic absence in middle school. In 2018, scoring 1 standard deviation higher on school-level academic challenge was associated with chronic absence rates that were 2.1 percentage points lower, on average.

***Interpreting Suspensions Results***

To create a suspensions variable with a more normal distribution, the study team calculated the natural log of the suspension index before running statistical analyses.[8] To help make coefficients from these models more interpretable, it is possible to convert the predicted percentage changes in the *log suspension index* to changes in the *suspension index*. This can be done by exponentiating the coefficient on the predictor of interest, subtracting one, and multiplying by 100. For example, for the regression of log suspension index on academic challenge, the coefficient on academic challenge was –0.33 for high school. Exponentiating –0.33 equals 0.72. Subtracting 1 from 0.72 equals –0.28 and multiplying by 100 equals –28. This represents a relative decrease in the suspension index of 28%.

Because relatively few students receive suspensions, on average, in any given school year, large *percentage changes* can translate into small *actual changes in the suspension index*. For example, the mean suspension index in sample high schools in 2018 was 8.8. A 1 standard deviation increase in the academic challenge scale was associated with a 28% lower suspension index in high schools. This 28% decrease translates to an expected decrease of 2.5 in the suspension index at the mean, from 8.8 to 6.3.

## Culture

Regarding academic outcomes, at the elementary level, the YouthTruth classroom culture scale demonstrated statistically significant evidence of concurrent validity for fifth-grade reading and math proficiency rates. At the secondary level, the school culture scale showed statistically significant evidence of concurrent validity for eighth-grade reading and math proficiency rates and ninth-grade retention rates. Furthermore, statistically significant findings supported the predictive validity of the school culture scale for eighth-grade reading proficiency rates.

Regarding behavioral outcomes, the school culture scale demonstrated statistically significant evidence of concurrent validity for middle school chronic absence rates and the middle and high school suspensions. Furthermore, the school culture scale demonstrated statistically significant evidence of predictive validity for high school suspensions.

***Elementary schools with stronger classroom cultures had higher same-year fifth-grade reading and math proficiency rates.*** In 2018, the classroom culture scale was associated with higher school proficiency rates in math and reading in fifth grade, after controlling for school demographics and state (Table 7). Scoring 1 standard deviation higher on the school culture scale was associated with a fifth-grade reading proficiency rate that was 4.0 percentage points higher and a fifth-grade math proficiency rate that was 5.9 percentage points higher, on average. At the elementary school level, the classroom culture scale was the only

[8] Recall that the suspension index is similar to the percentage of students receiving suspensions, but students with both in- and out-of-school suspensions are double counted.

survey scale examined that demonstrated statistically significant concurrent validity with reading or math proficiency rates.

***Middle schools with stronger school cultures had higher same-year eighth-grade reading and math proficiency rates and higher following-year eighth-grade reading proficiency rates.*** In 2018, the classroom culture scale was associated with higher school proficiency rates in reading and in eighth grade, after controlling for school demographics and state. In middle school, scoring 1 standard deviation higher on the school culture scale was linked to having eighth-grade reading and math proficiency rates that were both 5.6 percentage points higher, on average. Thus, the school culture scale demonstrated concurrent validity with respect to eighth-grade 8 reading and math proficiency rates.

Results also showed that, among middle schools, 2017 school culture scores were associated with increased 2018 reading proficiency rates, after controlling for school demographics and state. Each 1 standard deviation increase in middle school culture, as measured in 2017, was associated with a 4.7 percentage point increase in 2018 eighth-grade reading proficiency rate, on average.

### *Table 7. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes of interest on the YouthTruth culture scale by analysis type and school level*

| Outcome | Analysis type | Elementary | | Middle | | High | |
|---|---|---|---|---|---|---|---|
| **Academic outcomes** | | | | | | | |
| Reading proficiency rate | Concurrent | 3.95 | * | 5.56 | *** | 1.98 | |
| | Predictive | 2.37 | | 4.67 | *** | 2.86 | |
| Math proficiency rate | Concurrent | 5.88 | * | 5.58 | ** | 2.59 | |
| | Predictive | 3.22 | | 0.68 | | –1.35 | |
| Ninth-grade retention rate | Concurrent | NA | | NA | | –0.36 | * |
| | Predictive | NA | | NA | | NA | |
| **Behavioral outcomes** | | | | | | | |
| Chronic absence rate | Concurrent | -1.28 | | -3.60 | ** | -1.95 | |
| | Predictive | 1.08 | | 3.11 | | 0.67 | |
| Log suspension index | Concurrent | -0.18 | | -0.40 | ** | -0.40 | *** |
| | Predictive | -0.29 | | -0.15 | | -0.58 | *** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

*Note*. The elementary school scale focuses on classroom culture while the middle and high school scale focuses on school culture. Detailed information on variables and models are available in the Overview of Data and Methods section.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***School culture demonstrated concurrent validity with respect to ninth-grade retention rates.*** Among high schools that retained any ninth-grade students following the 2017–18 school year, those schools in which ninth graders reported higher levels of school culture during the 2017–18 school year had lower ninth-grade retention rates following the school year, after controlling for school demographics and state. A 1 standard deviation increase in average reported school culture by ninth graders was associated with an expected 0.36 percentage point decrease in the ninth-grade retention rate, on average.

***Middle schools with stronger school cultures showed lower chronic absence rates.*** Among middle schools with any chronically absent students in 2018, those schools reporting stronger school cultures had lower rates of chronic absence, after controlling for school demographics and state. A 1 standard deviation increase in the school culture scale was associated with an expected 3.6 percentage point decrease in the chronic absence rate, on average.

***Middle and high schools with stronger school cultures had lower suspensions.*** Among schools that suspended any students in 2018, suspensions decreased, on average, by 33% in middle schools and 33% in high schools with each 1 standard deviation increase in 2018 school culture, after controlling for school demographics and state. Consequently, the school culture scale exhibited concurrent validity with respect to suspensions in middle and high schools.

The school culture scale also displayed predictive validity with respect to suspensions in high school. In schools that had suspended any students in 2018, for each 1 standard deviation increase in the 2017 school culture scale, the 2018 suspensions decreased by 44% on average, after controlling for school demographics and state.

## Engagement

> The YouthTruth engagement scale demonstrated statistically significant evidence of concurrent validity for behavioral outcomes only: elementary school chronic absence rates, and the middle and high school suspensions. The student engagement scale also exhibited predictive validity for the high school suspensions.

***Elementary and middle schools with higher levels of student engagement had lower rates of chronic absence.*** In 2018, a 1 standard deviation increase in the engagement scale was associated with a 1.13 percentage point decrease in chronic absences in elementary schools and a 1.30 percentage point decrease in chronic absences in middle schools, on average, after controlling for school demographics and state (Table 8). Thus, the engagement scale demonstrated concurrent validity for elementary and middle school chronic absence rates. The engagement scale was the only YouthTruth student experience scale that was associated with decreased chronic absence rates in elementary school.

### *Table 8. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes of interest on the YouthTruth engagement scale by analysis type and school level*

| Outcome | Analysis type | Elementary | | Middle | | High | |
|---|---|---|---|---|---|---|---|
| **Academic outcomes** | | | | | | | |
| Reading proficiency rate | Concurrent | 1.46 | | 2.93 | | 0.60 | |
| | Predictive | 1.94 | | −0.22 | | −0.23 | |
| Math proficiency rate | Concurrent | 2.65 | | 1.99 | | −0.75 | |
| | Predictive | 1.05 | * | −2.22 | * | −3.01 | |
| Ninth-grade retention rate | Concurrent | NA | | NA | | -0.17 | |
| | Predictive | NA | | NA | | NA | |
| **Behavioral outcomes** | | | | | | | |
| Chronic absence rate | Concurrent | -1.13 | *** | -1.30 | *** | -0.33 | |
| | Predictive | -1.41 | | 3.79 | | 1.84 | |
| Log suspension index | Concurrent | -0.11 | | -0.33 | * | -0.37 | ** |
| | Predictive | -0.01 | | -0.13 | | -0.31 | ** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

*Note.* Detailed information on variables and models are available in the Overview of Data and Methods section.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***The engagement scale demonstrated concurrent validity for suspensions in both middle and high school, and predictive validity for suspensions in high school.*** Among schools that had suspended any student in 2018, a 1 standard deviation increase the engagement scale was associated with a 28% decrease in suspensions in middle school and 31% in high school, on average, after controlling for school demographics and state. In addition, in high schools, a 1 standard deviation increase in the engagement scale in 2017 was associated with a 27% decrease in suspensions, on average, after controlling for school demographics and state.

***The engagement scale did not show statistically significant evidence of concurrent validity or consistent evidence of predictive validity with respect to academic outcomes.*** Across school levels, no statistically significant relationships were detected between the engagement scale and same-year reading or math proficiency rates or ninth-grade retention rates, after controlling for school demographics and state. Predictive validity findings were mixed, with engagement measures in 2017 being positively related with 2018 fifth-grade math proficiency rates and negatively related with 2018 eighth-grade math proficiency rates, on average, after controlling for school demographics and state. No statistically significant findings were apparent between prior year engagement and reading proficiency rates or high school math proficiency rates, after controlling for school demographics and state.

## Relationships with Teachers

> With regard to academic outcomes, the YouthTruth relationships with teachers scale demonstrated statistically significant evidence of concurrent and predictive validity for eighth-grade reading proficiency rates.
>
> Regarding behavioral outcomes, the relationships with teachers scale showed concurrent validity for middle school chronic absence rates as well as middle and high school suspensions Additionally, the scale exhibited predictive validity for high school suspensions.
>
> Although the relationships with teachers scale was associated with key outcomes at the secondary level, associations between the scale and outcomes at the elementary level did not meet the threshold for statistical significance.

***Middle schools with higher scores on the relationships with teachers scale in 2017 or 2018 had higher 2018 reading proficiency rates.*** The more that students in a middle school reported receiving support and personal attention from their teachers (that is, the higher the average relationships with teachers scale score), the higher the school's same- and following-year school eighth-grade reading proficiency rates were, on average (Table 9). On average, each 1 standard deviation increase in the 2018 middle school relationships with teachers scale was associated with an increase of 3.9 percentage points in the 2018 eighth-grade reading proficiency rate, after controlling for school demographics and state. Similarly, each 1 standard deviation increase in the 2017 middle school relationship scale was associated with an increase of 2.9 percentage points in the 2018 eighth-grade reading proficiency rate, on average after controlling for school demographics and state. The relationships with teachers scale demonstrated concurrent and predictive validity with respect to eighth-grade reading proficiency rates.

***In 2018, chronic absence rates were lower in middle schools in which students received more support and attention from their teachers, according to the relationships with teachers scale.*** In middle schools reporting any chronic absences, chronic absence rates were about 2.2 percentage points lower, on average, for each 1 standard deviation increase in the relationships with teachers scale, after controlling for school demographics and state.

***Middle and high schools with higher average scores on the relationships with teachers scale had lower suspensions.*** In middle schools reporting any student suspensions, each 1 standard deviation increase in the 2018 school relationships with teachers scale was associated with an expected 26% reduction in 2018 suspensions, on average, after controlling for school demographics and state. In high schools, this reduction was 24%. Furthermore, high schools exhibited predictive validity with respect to suspensions. Each 1 standard deviation increase in the 2017 relationships with teachers scale was associated with an

average decrease of 34% in 2018 suspensions, on average, after controlling for school demographics and state.

### *Table 9. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes on the YouthTruth relationships scale by analysis type and school level*

| Outcome | Analysis type | Elementary | Middle | | High | |
|---|---|---|---|---|---|---|
| **Academic outcomes** | | | | | | |
| Reading proficiency rate | Concurrent | 0.57 | 3.94 | ** | 1.69 | |
| | Predictive | 1.56 | 2.88 | *** | 3.08 | |
| Math proficiency rate | Concurrent | 1.24 | 3.49 | | 0.38 | |
| | Predictive | 0.79 | 0.29 | | –1.53 | |
| Ninth-grade retention rate | Concurrent | NA | NA | | -0.28 | |
| | Predictive | NA | NA | | NA | |
| **Behavioral outcomes** | | | | | | |
| Chronic absence rate | Concurrent | -1.37 | -2.21 | ** | -0.14 | |
| | Predictive | 0.59 | 2.18 | | -0.35 | |
| Log suspension index | Concurrent | -0.12 | -0.30 | ** | -0.27 | *** |
| | Predictive | -0.10 | -0.16 | | -0.42 | *** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

*Note.* Detailed information on variables and models are available in the Overview of Data and Methods section.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***At the elementary school level, the relationships with teachers scale was not statistically significantly related to any outcome examined.*** Although the relationships with teachers scale was associated with key outcomes at the middle and high school levels, associations between the elementary relationships with teachers scale and outcomes at the elementary level did not meet the threshold for statistical significance, despite the relationships being in the predicted direction. Importantly, the items composing the elementary school relationships with teachers scale differ substantively from those making up the secondary school relationships with teachers scale and have a 3- rather than 5-point response scale. Although most YouthTruth student experience scales saw fewer statistically significant relationships at the elementary level than at the middle or high school levels, the relationships with teachers scale was the only one that had no statistically significant relationships at the elementary level.

## Belonging and Peer Collaboration

> The belonging and peer collaboration scale, which was collected for middle and high school students, exhibited statistically significant evidence of concurrent validity for several academic outcomes: eighth-grade reading proficiency rates, high school reading proficiency rates, and high school math proficiency rates. The scale also showed statistically significant evidence of predictive validity for eighth-grade and high school reading proficiency rates.
>
> Regarding behavioral outcomes, the belonging and peer collaboration scale demonstrated concurrent and predictive validity for high school chronic absence rates and middle school suspensions.

Belonging and peer collaboration was the only YouthTruth student experience scale to demonstrate statistically significant evidence of concurrent validity for high school math proficiency rates and high school chronic absence rates. It was also the only scale to demonstrate statistically significant predictive validity for middle school suspensions. Further, the scale had the largest apparent effects at the high school level on reading proficiency rates, math proficiency rates, and chronic absence rates (Table 10).

***The belonging and peer collaboration scale demonstrated concurrent and predictive validity with respect to eighth-grade and high school reading proficiency rates.*** Higher levels of belonging and peer collaboration as measured by the belonging and peer collaboration scale were associated with higher eighth-grade and high school reading proficiency rates in 2018, after controlling for school demographics and state. Specifically, scoring 1 standard deviation higher on the belonging and peer collaboration scale was associated with an eighth-grade reading proficiency rate that was 5.3 percentage points higher and a high school reading proficiency rate that was 7.7 percentage points higher, on average.

Higher levels of belonging and peer collaboration in 2017 were associated with higher 2018 eighth-grade and high school reading proficiency rates, after controlling for school demographics and state. Specifically, scoring 1 standard deviation higher on the 2017 belonging and peer collaboration scale was associated with a 2018 eighth-grade reading proficiency rate that was 3.0 percentage points higher and a 2018 high school reading proficiency rate that was 5.9 percentage points higher, on average.

***The belonging and peer collaboration scale was positively linked with same-year high school math proficiency rates.*** After controlling for school demographics and state, a 1 standard deviation increase in belonging and peer collaboration was associated with a high school math proficiency rate that was 6.3 percentage points higher, on average. The belonging and peer collaboration scale was the only YouthTruth measure investigated that showed a statistically significant same-year relationship with math proficiency rates in high school.

### *Table 10. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes on the YouthTruth belonging and peer collaboration scale by analysis type and school level*

| Outcome | Analysis type | Middle | | High | |
|---|---|---|---|---|---|
| **Academic outcomes** | | | | | |
| Reading proficiency rate | Concurrent | 5.34 | ** | 7.70 | *** |
| | Predictive | 3.03 | * | 5.85 | ** |
| Math proficiency rate | Concurrent | 7.03 | | 6.33 | *** |
| | Predictive | 0.30 | | 1.89 | |
| Ninth-grade retention rate | Concurrent | NA | | 0.07 | |
| | Predictive | NA | | NA | |
| **Behavioral Outcomes** | | | | | |
| Chronic absence rate | Concurrent | -2.60 | | -2.16 | ** |
| | Predictive | 3.13 | | -2.78 | ** |
| Log suspension index | Concurrent | -0.36 | ** | -0.12 | |
| | Predictive | -0.12 | * | -0.06 | |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

Note. The belonging and peer collaboration scale applied to students in middle and high schools only. Detailed information on variables and models are available in the Overview of Data and Methods section.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***In 2018, chronic absence rates were lower in high schools that had higher belonging and peer collaboration scores on average.*** On average, in sample high schools in 2018, chronic absence rates were 2.2 percentage points lower for every 1 standard deviation increase in the belonging and peer collaboration scale.

***Middle schools with a higher level of belonging and peer collaboration saw better same-year and following-year student discipline, as measured by suspensions.*** On average in 2018, scoring 1 standard deviation higher on the belonging and peer collaboration scale was associated with a predicted 30% decrease in the 2018 school average suspensions, after controlling for school demographics and state. Additionally, scoring 1 standard deviation higher on the belonging and peer collaboration scale in 2017 was associated with a predicted 11% decrease in the 2018 school average suspensions, after controlling for school demographics and state, on average. Jointly, these findings provide statistically significant evidence that the belonging and peer collaboration scale has concurrent and predictive validity with respect to the middle school suspensions.

## College and Career Readiness

> The college and career readiness scale, which was collected only from high school students, demonstrated concurrent and predictive validity for suspensions, as well as a negative concurrent validity with reading proficiency rates.

***High schools with higher scores on the college and career readiness scale in 2017 or 2018 saw lower levels of suspensions in 2018.*** Among high schools in which any students were suspended, a 1 standard deviation increase the college and career readiness scale in 2018 was associated with a 27% decrease in suspensions in 2018 (Table 11). Similarly, a 1 standard deviation increase in the college and career readiness scale in 2017 was associated with a 33% decrease in suspensions in 2018.

***Table 11. Estimated coefficients of ordinary least squares regressions of academic and behavioral outcomes on the YouthTruth college and career readiness scale by analysis type and school level***

| Outcome | Analysis type | High | |
|---|---|---|---|
| **Academic outcomes** | | | |
| Reading proficiency rate | Concurrent | −2.68 | * |
| | Predictive | −0.19 | |
| Math proficiency rate | Concurrent | −3.64 | |
| | Predictive | −4.34 | |
| Ninth-grade retention rate | Concurrent | −0.23 | |
| | Predictive | NA | |
| **Behavioral outcomes** | | | |
| Chronic absence rate | Concurrent | -0.57 | |
| | Predictive | -0.25 | |
| Log suspension index | Concurrent | −0.32 | * |
| | Predictive | −0.40 | ** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

NA = not applicable.

*Note*. Detailed information on variables and models are available in the Overview of Data and Methods section.

*Source*. SRI Education analyses of data from YouthTruth and from the U.S. Department of Education's Civil Rights Data Collection and restricted-use ED*Facts* data file (Results for State Assessments Achievement in Reading Language Arts and Mathematics 2016–2018).

***Schools with higher levels of school support for college and careers had lower high school reading proficiency rates.*** The college and career readiness scale demonstrated a statistically significant relationship with an outcome of interest in an unexpected direction. A negative relationship was evident between 2018 levels of school support for college and career readiness and 2018 high school reading performance. In other words,

higher levels of support for college and career readiness in 2018 were associated with lower high school reading proficiency rates.[9]

# Chapter IV: Discussion

This study provides clear school-level evidence of the concurrent and predictive validity of the YouthTruth student experience scales. The study has made a substantial contribution to the literature by showing that each YouthTruth student experience scale demonstrates both concurrent validity for one or more key academic and behavioral outcomes in at least one school level. The same holds for predictive validity. These findings support the use of the YouthTruth student experience scales in school improvement efforts. School and district staff can be confident that school-level results on the YouthTruth student survey are associated with key school-level educational outcomes of interest—reading and math proficiency, ninth-grade retention, chronic absences, and suspensions. Furthermore, this study has begun to demonstrate which student experience scales to focus on to boost particular student outcomes.

## Limitations

Several factors may have limited the study team's ability to fully detect associations between the scales and school-level outcomes. Together, these factors mean that additional associations between the YouthTruth student experience scales and key outcomes may exist beyond those presented in this report.

***Sample Size.*** The small number of schools in each sample, particularly the ninth-grade retention and some of the chronic absence samples, mean that these analyses have low statistical power to detect true underlying relationships between the student experience scales and outcomes of interest.

***Outcomes.*** This study did not examine the relationship between YouthTruth student experience scales and other important student outcomes, including students' social and emotional well- being and longer term outcomes like college enrollment and persistence.

***Aggregation Bias.*** Aggregating data from lower to higher units of analysis (e.g., from student to schools) can make it harder to detect associations between variables because the aggregation reduces variation. The study team had access to only school- or grade-level aggregated data on both YouthTruth student experience scales and student outcomes, so all results are based on

[9] This finding appears counterintuitive as higher skills in English language arts are typically associated with increased college and career readiness among students. One possible explanation (among many) for this finding is that schools with lower levels of student reading performance may be more likely to emphasize career readiness for their students, which would raise the overall college and career readiness scale score. Alternatively, this result could be due to chance, given the large numbers of analyses run.

between-school variation in predictors and outcomes of interest. Further disaggregating data to the student level might enable the detection of additional underlying school-level relationships.

***Predictor and Outcome Alignment.*** Two outcomes, reading and math proficiency rates, were limited to a single grade (fifth grade in elementary schools, eighth grade in middle schools, and a single grade in high school that varied by state). In contrast, the study's predictors of interest were drawn from students enrolled in all grades within each school, except for those used in the ninth-grade retention analyses. Findings for reading and math proficiency rates may have been different if the match between outcomes and predictors had been more exact.

Finally, the school-level associations between the YouthTruth student experience scales and student academic and behavioral outcomes are correlational, not causal. The associations the study team found do not mean that an increase in one of these scales is solely responsible for an improved outcome. For example, a statistically significant positive relationship between the belonging and peer collaboration scale and high school reading proficiency rates does not mean that increased belonging and peer collaboration caused increased high school reading proficiency rates. Rather, higher reading proficiency rates could lead to increased belonging and peer collaboration, or a third factor or a constellation of additional factors could be associated with both belonging and peer collaboration and reading proficiency. Although the study team controlled for a number of school compositional variables—such as gender, race, school size, and economic disadvantage—there may be differences among schools that the team could not account for in analyses, and these could underly some of the detected associations.

## Potential Future Research

Future research could examine additional student- or school-level outcomes of interest. For example, additional outcomes may include academic outcomes in other subject areas, such as science, social studies, or history, as well as measures of educational progress and attainment. Additional behavioral outcomes ripe for investigation include bullying, class participation, paying attention in class, and treating others with respect. The relationships among the YouthTruth student experience scales and measures of students' social and emotional skills—including motivation, self-efficacy, social awareness, and self-regulation—also merit exploration. Student mental health indicators such as anxiety or depression are important outcomes worthy of future study. In addition, future analyses of the linkages between YouthTruth student experience scales and reading and math performance could draw on newly available data from the Stanford Education Data Archive to examine student achievement data across grades and states.

Further, future studies drawing on student-level data would be more sensitive in detecting school-level concurrent and predictive validity of the YouthTruth student experience scales. Analyzing student-level data via multilevel models would enable investigations into individual- and school-level effects of the student experience scales on outcomes of interest. Additionally,

analyses of multiple years of student-level data would facilitate a better understanding of how changes in student experiences relate to changes in outcomes of interest. Compared with school-level data, student-level data has more variation in predictors and in outcomes and would enable analyses that do not suffer from aggregation bias. And student-level data would generate substantially more power to detect effects: consider that a single school may have more students enrolled than the school-level sample sizes in this study.

Given the results from this study, researchers investigating causal effects of education interventions may wish to consider using the YouthTruth student experience scales as measures of interim or ultimate outcomes for their studies. For example, if an intervention aimed to reduce chronic absenteeism in a high school, researchers may wish to measure belonging and peer collaboration using the corresponding YouthTruth student experience scale, to investigate whether any potential impacts of the intervention on student retention were moderated by changes to belonging and peer collaboration within the school. Likewise, researchers designing future studies of the impact of interventions on school environment or school climate outcomes may consider using YouthTruth student experience scales as baseline measures and outcomes of interest.

# References

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. Review of Educational Research, 87(2), 425–469. https://doi.org/10.3102/0034654316669821

National Center for Education Statistics. (2020). Percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and state or jurisdiction: Fall 2000 and fall 2018 (Table 203.70). Digest of education statistics. https://nces.ed.gov/programs/digest/d20/tables/dt20_203.70.asp

YouthTruth. (2018). Design & methodology: Student survey. https://youthtruthsurvey.org/wp-content/uploads/2020/02/YouthTruth-Design-and-Methodology-Student.pdf