

**Working Paper:  
Measurement Properties of Student Social-Emotional Competency and School Culture-  
Climate Surveys in the NewSchools Invent Cohort**

Katie Buckley  
Sushmita Subedi  
Sara Krachman  
*Transforming Education*

Jason Atwood  
*NewSchools Venture Fund*

with  
*Education Analytics*

October 9<sup>th</sup>, 2018

This paper was produced as part of a multi-year research partnership between Transforming Education and NewSchools Venture Fund. Please direct all feedback to Katie Buckley: [katie@transformingeducation.org](mailto:katie@transformingeducation.org). We are grateful for the analytic support of Education Analytics (EdAnalytics; <http://edanalytics.org>), a leading measurement and evaluation organization partnering with the CORE districts to conduct analytics and evaluation of their continuous improvement system. For the purposes of this project, EdAnalytics conducted several analyses on the measurement properties of the surveys and reviewed this report for technical accuracy. We also thank Andrew Ho, Susan Lyons, Sarah Phillips and Katerina Schenke for their helpful reviews of earlier drafts.

## **Measurement Properties of Student Social-Emotional Competency and School Culture-Climate Surveys in the NewSchools Invent Cohort**

### **Abstract**

This study describes the measurement properties of two curated surveys, consisting of previously developed scales, assessing student perceptions of their social-emotional competencies (SEC) and the school culture/climate (CC). The surveys were administered through the Transforming Education/NewSchools Venture Fund partnership that worked with the NewSchools Invent Cohort, a network of 23 schools across 11 states, to expand the definition of student success. The analytic sample for this study includes more than 3,000 students in grades 4 through 12 at 18 schools surveyed in the 2016-17 school year. Using classical test theory as our conceptual and analytic lens, we conducted tests of internal consistency, item discrimination, omit rates, differential item functioning, and exploratory and confirmatory factor analysis. Our findings provide initial evidence of the validity of the survey scores for measuring student perceptions of their SEC and the school CC in the NewSchools Invent schools, suggesting that the surveys are suitable for school and cohort-level decision-making related to students' social emotional competencies and experiences within the school.

## Introduction

Research has demonstrated the importance of intra- and interpersonal social-emotional competencies (SEC) for positive academic and life outcomes. Skills such as self-regulation and social competence and motivational orientations such as growth mindset predict outcomes ranging from high school and college completion to employability and earnings, as well as physical and mental health (see reviews in Gabrieli, Ansel, & Krachman, 2015; Jones, Greenberg, & Crowley, 2015; Moffit, et al., 2011). We also know that these skills and mindsets do not develop in a vacuum. Aspects of a school’s culture/climate (CC)—for example, students’ perception of sense of belonging and teacher-student relationships—can foster students’ academic and SEC development (Anderman, 2003; Berg et al., 2017; Konishi et al., 2010). Students with strong SEC are more likely to build positive relationships with peers and adults, participate in classroom activities, and engage in learning (Jones & Kahn, 2017). In turn, schools that are intentional about combining rigorous academics with social-emotional development produce deeper, longer-term student learning (Farrington et al., 2012; Jones & Kahn, 2017).

To rigorously examine the interconnections between academic outcomes, social-emotional competency development and school environment, the field needs valid and reliable measures of student SEC and the school CC across grade levels and demographic subgroups. In partnership with NewSchools Venture Fund, Transforming Education administered a set of pre-existing self-report SEC scales and school CC scales as two new comprehensive student surveys.<sup>1</sup> In this study, we examine the measurement properties of the SEC and CC surveys,

---

<sup>1</sup> Transforming Education (TransformEd) and NewSchools Venture Fund (NewSchools) have partnered together in an effort to help schools in the NewSchools Invent cohort address a broader range of factors that contribute to students’ long-term success. This three-year partnership is focused on achieving four key goals: (1) Provide actionable data and research to help school leaders expand the definition of student success and to improve outcomes for students on a range of indicators that relate to long-term success; (2) Provide support (e.g., resources,

based on their administration in and use with the NewSchools Invent cohort. While there exists prior evidence of the reliability and validity for each individual SEC and CC scale, the validity of the new surveys has not been examined. We recognize that validation is an ongoing process of accumulating multiple sources of evidence to support the appropriateness of the decisions and inferences being made with the instrument (Kane, 2006; 2013; Messick, 1989). In this study, we examine the measurement properties of the surveys to provide evidence on a key aspect of validation – how well the measures were designed; that is, the extent to which the items provide consistent and new information about the underlying constructs being assessed, the extent to which the items are interpreted comparably across student subgroups, and the extent to which scales for each survey are measuring unique constructs. To conduct this study, we partnered with Education Analytics, an organization comprised of leading measurement experts who serve as the analytic partner to the CORE Districts. The set of analyses we include in this report were informed by analyses that EdAnalytics conducted on the CORE Districts’ SEC survey (see Meyer, Wang & Rice, 2018).<sup>2</sup>

## **Background**

There are different methods of assessing student SEC and school CC; student self-reports are one method that can provide insight into student perceptions of skills, mindsets, and experiences. Social and cognitive psychology literature suggests that self-reported questionnaires provide a good medium for respondents to communicate their true perceptions (Krosnick, 1999).

---

connections, ideas, etc.) to help school leaders change practices based on research and data; (3) Provide data to NewSchools to help them understand their portfolio on a variety of metrics so they can identify trends and inform their board, investment partners, and funders; and (4) Contribute to the broader national dialogue in the field about how to expand the definition of student success in research, policy and practice.

<sup>2</sup> The CORE districts are a collaborative of eight California school districts that have jointly developed a system of school accountability and continuous improvement that includes measures of social-emotional skills based on student self-reports.

In particular, self-report questionnaires “are arguably better suited than any other measure for assessing internal psychological states like feelings of belonging” (Duckworth & Yeager, 2015, p. 5). Furthermore, numerous studies have shown that results from self-report questionnaires are predictive of objectively-measured outcomes (Fredricks et al., 2011; Hanson, & Kim, 2007; Squires et al., 2011; West, Buckley, Krachman, & Bookman, 2017). On a practical level, self-report surveys are an easy and low-cost way of obtaining information about individual perceptions, attitudes, and behavior.

While student self-reports provide a practical and scalable solution for assessing student social-emotional competencies and illuminating student perceptions, researchers have raised questions about potential biases inherent in the self-reports (Duckworth & Yeager, 2015). West et al. (2017) recommended taking these concerns seriously but also emphasized the fact that self-reports provide a tremendous learning opportunity for the field. Further, a recent report by Policy Analysis for California Education (PACE) at Stanford University found promising evidence for the validity of student self-reports of social-emotional competencies administered in the CORE districts as part of CORE’s continuous improvement system (Gehlbach & Hough, 2018).

Addressing concerns about self-report questionnaires requires continuous testing and validation of existing measures to ensure their proper use. In order to better serve students and to help schools make effective data-informed decisions, by, for example, linking formative SEC assessments to specific instructional practices, it is imperative that we use measures of student SEC and school CC that are able to assess such constructs equally well across age, gender, race/ethnicity, and socio-economic groups (Phillips & Rowley, 2015). Furthermore, validation of SEC and CC measures is an important prerequisite to examining theoretical relationships among student academic performance, student SEC, and school CC. In this study, we examine the

measurement properties of the SEC and CC survey items and scales using classical test theory. In line with Meyers et al., (2018), we provide a thorough review of properties of the items and scales to investigate the extent to which surveys can be used reliably for school- and cohort-level decision-making about how to select and implement strategies, resources, and policies that improve student outcomes.

## **Data**

### **Instrumentation**

In the summer of 2016, TransformEd partnered with school leaders who were members of the NewSchools Invent cohort of innovative schools. The goal of this partnership was to help school leaders expand their definition of student success through the collection and analysis of data on student social-emotional competencies and school culture-climate factors. The first step toward achieving this goal was to determine which SECs and CC factors should be assessed across the cohort of schools based on common measures. From interviews with school leaders in the NewSchools Invent cohort, TransformEd generated a list of more than 60 constructs that school leaders believed to be most important to students' long-term success. TransformEd then filtered this list through its "3Ms framework," narrowing the list to constructs that are *meaningful* (have an impact on long-term student outcomes), *measurable* (can be assessed in a school setting), and *malleable* (can be developed in a school setting), based on existing literature from the fields of economics, psychology, human development, and education.<sup>3</sup> The final set of constructs include seven social-emotional competencies (e.g., self-regulation and growth mindset) and seven school culture-climate factors (e.g., school safety and student-teacher

---

<sup>3</sup> See <https://www.transformingeducation.org/wp-content/uploads/2017/06/Tool-3.pdf>

relationships).

TransformEd scanned the field to identify a set of pre-existing scales (i.e., validated and free to use for non-commercial research purposes) that measured student perceptions of these 14 prioritized constructs. The SEC scales were drawn from leading researchers and are used in surveys administered by the CORE Districts, the Boston Charter Research Collaborative (BCRC), Washoe County School District, and the National Assessment of Education Progress (NAEP).<sup>4</sup> The school CC scales were drawn from Panorama Education’s school climate survey and the U.S. Department of Education’s ED School Climate Survey (EDSCLS).<sup>5</sup> TransformEd assembled the scales into a new SEC survey and CC survey to administer to students in schools that were part of the NewSchools Invent cohort in the 2016-17 school year. See Appendix A for the full set of scales, items, and sources.

**Student-reported SEC survey.** A 39-item SEC survey was used to measure student perceptions of seven different competencies.<sup>6</sup> For each item, students selected one of five Likert-type options (*strongly disagree* to *strongly agree*).

- Curiosity (6 items, e.g., “I like activities that challenge my thinking abilities”)<sup>7</sup>
- Growth mindset (4 items, e.g., “Challenging myself won’t make me any smarter”)<sup>8</sup>

---

<sup>4</sup> For further background on our partnership with the CORE Districts and BCRC, please see: <https://www.transformingeducation.org/our-work/with-schools-and-systems/core-districts-partnerships/> and <https://www.transformingeducation.org/our-work/with-schools-and-systems/bcrc-partnership/>. For more information on Washoe County School District’s SEL work, please see: <http://www.wcsddata.net/data-topics/sel/>. For further information on NAEP’s SEC scales, please see <https://nces.ed.gov/nationsreportcard/tdw/instruments/noncog.aspx>

<sup>5</sup> For further background on the school climate scales, please visit: <https://www.panoramaed.com/school-climate-survey> and [https://nces.ed.gov/surveys/edscls/pdf/EDSCLS\\_Student\\_Questionnaire\\_English.pdf](https://nces.ed.gov/surveys/edscls/pdf/EDSCLS_Student_Questionnaire_English.pdf).

<sup>6</sup> The full list of SEC items for each scale are listed in Table 5. See Tables A2 and A3 in the appendix for overall and subgroup-level means and standard deviations of each SEC and CC competency.

<sup>7</sup> The curiosity scale was drawn from the National Assessment of Educational Progress (NAEP). NAEP refers to this construct as “desire for learning.”

<sup>8</sup> The growth mindset scale was developed by Camille Farrington of the University of Chicago Consortium on Chicago School Research (CCSR), building upon the work of Carol Dweck of Stanford University. Items in this scale are reverse-coded.

- Perseverance (4 items, e.g., “I finish whatever I begin”)<sup>9</sup>
- Self-awareness (4 items, e.g., “Please let us know how easy or difficult each of the following are for you: “Knowing what my strengths are”)<sup>10</sup>
- Self-efficacy (4 items, e.g., “I can earn an A in my classes”)<sup>11</sup>
- Self-regulation (9 items, e.g., “I came to class prepared”)<sup>12</sup>
- Social awareness (8 items, e.g., “I listened carefully to other people’s points of view”)<sup>13</sup>

**Student-reported culture-climate survey.** The 36-item culture-climate survey was used to measure student perceptions of the school CC across seven domains. For each item, students select one of five Likert-type options, (*strongly disagree* to *strongly agree*).<sup>14</sup>

- Cultural and linguistic competence (5 items, e.g., “Boys and girls are treated equally well”)<sup>15</sup>

---

<sup>9</sup> The perseverance scale was developed by Angela Duckworth of the University of Pennsylvania. Duckworth refers to this construct as “grit.” The original scale consisted of the items included in the SEC survey, however it has since been modified to include two factors: consistency of interest and perseverance of effort. Our scale aligns closely with the items that fall under perseverance of effort factor (see [https://www.dropbox.com/s/rn5wo3y0iis0qtf/8-item%20Grit%20Scale\\_Child%20Adapted%20Version\\_4.pdf?dl=0](https://www.dropbox.com/s/rn5wo3y0iis0qtf/8-item%20Grit%20Scale_Child%20Adapted%20Version_4.pdf?dl=0)).

<sup>10</sup> The self-awareness scale was developed by Washoe County, American Institutes for Research (AIR) and the Collaborative for Academic, Social, and Emotional Learning (CASEL) as part of the Collaborative Districts Initiative (CDI).

<sup>11</sup> The self-efficacy scale was developed by Camille Farrington of the University of Chicago Consortium on School Research (CCSR).

<sup>12</sup> The self-regulation scale was developed by Angela Duckworth of the University of Pennsylvania and adapted by Clancy Blair of New York University (Duckworth, 2008; Blair & Diamond, 2008). Duckworth and Blair refer to this construct as “self-control” (Duckworth, Gendler, & Gross, 2016). While there are substantive differences in the underlying constructs of self-control and self-regulation (see for example, <https://www.psychologytoday.com/us/blog/self-reg/201607/self-reg-self-regulation-vs-self-control>), we refer to the scale as self-regulation based on feedback from the school leaders taking part in this study.

<sup>13</sup> The social awareness scale was developed by CASEL and AIR and adapted by Hunter Gehlbach for use by the CORE Districts.

<sup>14</sup> Three of the SEC and CC scales are reverse coded: the growth mindset scale (due to negatively framed questions), the cultural and linguistic competence scale (with response options listed as strongly agree to strongly disagree and school safety scale (with questions asking about lack of school safety within a school).

<sup>15</sup> The cultural and linguistic competence scale was drawn from the EDSCLS. It is based on a 4-point Likert Scale. For the purposes of our survey, we converted it to a 5-point Likert scale to be consistent with the other six CC scales drawn from Panorama.

- Learning strategies (5 items, e.g., “Before you start on a challenging project, how often do you think about the best way to approach the project?”)
- Rigorous expectations (5 items, e.g., “How often do your teachers make you explain your answers?”)<sup>16</sup>
- School safety (6 items, e.g., “How often do you worry about violence at your school?”)
- Sense of belonging (5 items, e.g., “How connected do you feel to the adults at your school?”)
- Student engagement (5 items, e.g., “How excited are you about going to your classes?”)
- Teacher-student relationships (5 items, e.g., “How many of your teachers are respectful towards you?”)

### **Administration**

The surveys were administered through an online platform, managed by Panorama Education, to students in grades 4-12 across 18 schools in the NewSchools Invent cohort in 2016-17. See Appendix B for summary statistics for each of the 14 competencies from the Spring 2017 survey administration. Students took the survey in the first 4-6 weeks of the fall semester school and again in the spring semester during the last 4-6 weeks of school. Students were asked to complete both surveys in one sitting, which generally took approximately 45 minutes. While there was no official protocol on where and when students took the surveys, most schools chose to administer surveys during an advisory or homeroom period to minimize the impact on academic instructional time.

---

<sup>16</sup> The learning strategies scale, rigorous expectations scale, school safety scale, sense of belonging scale, student engagement scale, and teacher-student relationship scale were developed by Panorama Education (see <https://www.panoramaed.com/school-climate-survey>).

## Analytic Sample

Table 1 summarizes the demographic characteristics of our analytic sample based on student roster results. Note that our analytic sample excludes students from one eligible school that chose not to administer the surveys, as well as students in grades 4-12 who were eligible to take the surveys but chose not to or were absent on the days the surveys were administered. We also excluded students who took the survey in Spanish due to small sample sizes.<sup>17</sup> Our final analytic samples include all students who answered at least one item on each survey in each time period.

**Table 1.** Demographic characteristics of students in the analytic sample

Characteristic	Fall 2016 Sample with Student Survey Data		Spring 2017 Sample with Student Survey Data	
	Student SEC	Student CC	Student SEC	Student CC
Number of schools	18	18	18	18
Number of students (Grades 4 through 12)	n=3,222	n=3,274	n=3,021	n=3,086
Female	48%	48%	49%	50%
Latino	53%	52%	53%	52%
Black or African American	20%	20%	18%	18%
Asian	3%	3%	3%	3%
Two or More Races	1%	1%	1%	1%
White	19%	20%	21%	22%
Other	0.53%	0.52%	0.53%	0.49%
Middle Eastern North African(MENA)	3%	3%	3%	3%
ELL	17%	18%	18%	18%
FRPL	65%	65%	64%	64%
SPED	8%	8%	8%	9%

<sup>17</sup> For the Fall analytic sample, we exclude sixteen students (0.49%) who took the SEC survey in Spanish, and eighteen (0.55%) who took the CC survey in Spanish. For the Spring analytic sample, we excluded seven students (0.23%) who took the SEC survey in Spanish and six students (0.19%) who took the CC survey in Spanish.

## **Methods**

We used several diagnostic methods in line with classical test theory to examine the measurement properties of the SEC and CC survey items and scales (Hambleton & Jones, 1993; Jiang, Wang, & Weiss, 2016). We conducted most of our psychometric analyses based on classical test theory instead of item response theory (IRT) due to sample size limitations at the subgroup level (student demographics and grade level). Studies suggest that a sample size of 400 or greater is required to obtain estimates with useful standard errors (He & Wheadon, 2013; Jiang, Wang, & Weiss, 2016).

To assess the quality of SEC and CC items and scales, we first examined the internal consistency of each SEC and CC scale by conducting tests of reliability using Cronbach's alpha coefficient of each scale. We then conducted three psychometric diagnostics using the statistical software R and Stata. First, we examined item response spread to identify the proportion of students that endorsed a particular item. The purpose of this analysis was to determine whether the items included in our scales are able to capture the full range of student perceptions. Second, we examined item discrimination to determine whether items adequately differentiate between students with high perceptions (of SEC or CC) and low perceptions (of SEC or CC). This type of analysis can illuminate the extent to which the items are cohesively measuring the underlying construct. Third, we analyzed omit rates to identify items that were systematically skipped by students. Analyzing omit rates is important because if items within a scale are systematically skipped by a considerable number of students, it could bias aggregate responses of the scale (Phillips, Reddy, & Durning, 2016).

In addition, we conducted differential item functioning (DIF) analysis on the survey items to identify whether there are statistically significant differences in student performance on

an item between the reference group (e.g., male) and the focal group (e.g., female) among students with similar levels of self-perceived SEC competencies or similar perceptions of school CC (Meyers et al., 2018).

DIF is an important analytic tool because it provides insight into the extent to which the survey is capable of producing comparable responses across student subgroups (Guerra & Jagers, 1998; Rogers, 2014). This analysis is particularly relevant for the SEC and CC surveys as its participants are a diverse group of students based on their race/ethnicity, gender, and grade level. As such, it is possible that students from a particular subgroup may answer an SEC or CC item in a particular way not because of their true perception, but because of societal factors that influence their beliefs in how they should respond (i.e., social desirability bias) or because of difficulty in understanding a particular word or phrase. Jeong & Lee (2016) provided an example of differences in Likert-type responses on a self-confidence item between respondents from a country where humility is encouraged versus a country where self-assurance is encouraged. They used DIF to identify such unequal responding patterns among groups. While DIF does not necessarily indicate bias, it may simply highlight important differences across subgroups in students' perceptions or experiences of a particular aspect of the school culture-climate. Identifying and further investigating such items is an important step toward creating a survey that produces comparable scores across all student subgroups (Raju et al., 2002).

Lastly, we performed exploratory and confirmatory factor analyses to assess whether the scales within each survey were measuring distinct constructs. Exploratory factor analysis (EFA) assumes no underlying factor structure and therefore allows us to freely estimate the factor structure across all items within a survey. Confirmatory factor analysis (CFA) allows us to impose theoretically expected factor structures to test the hypothesis that the desired relationship

between observed survey items and their underlying latent constructs exists. With the SEC and CC surveys, we impose a 7-factor structure since each includes seven distinct scales. If, through EFA and CFA, two scales are found to be measuring the same underlying construct, it suggests that the two scales can be combined or that one of the scales can be removed.

## Findings

### Internal consistency

We assess internal consistency by examining Cronbach's alpha coefficient for each of the survey scales. Cronbach's alpha coefficient ( $\alpha$ ) assesses the extent to which items within a scale measure the same construct. In general, a higher Cronbach's alpha coefficient indicates greater consistency of item responses.<sup>18</sup> Alpha coefficients of 0.60 to 0.70 are considered common in survey research (Lamb et al., 2012; Ravens-Sieberer et al., 2008).

Generally, we find that the scales in the SEC and CC surveys demonstrate a high degree of internal consistency across all students in the NewSchools Invent cohort. For SEC scales, alpha coefficients range from 0.70 to 0.92 for grades 4 through 12. For CC scales, alpha coefficients range from 0.66 to 0.91 for grades 4 through 12. See Appendix C for grade-level alpha coefficients for both SEC and CC surveys.

### Item response spread

The item response spread statistic indicates the degree to which students endorse an item. It is calculated as the proportion of the maximum obtainable score and ranges between 0 and 1. If an item is endorsed by a large proportion of students (i.e., most students select a 4 or a 5 on a 5-point Likert scale) or an item is endorsed by a small proportion of students (i.e., most students

---

<sup>18</sup> Cronbach's alpha coefficient provides an imperfect measure of internal reliability, since it is affected by the number of items included in the scale, and high or perfect correlations can suggest item redundancy and/or a narrow measure of the underlying construct (see Tavakol & Rennick, 2011).

select a 1 or a 2 on a 5-point Likert scale), this can be an indicator that the item is either too “easy” or too “difficult” to provide meaningful information about the relevant construct being measured. For the purposes of this study, we use the threshold of 0.30 to 0.80 for a constructed response item (see Meyer et al., 2018), whereby an item is deemed too “difficult” if the average endorsement is below 30 percent and too “easy” if average endorsement is above 80 percent.

While the terms “hard” and “easy” are generally used for cognitive items with correct or incorrect answers, the meaning can be less intuitive for items assessing SEC and CC constructs, particularly those assessing non-skill factors. That said, in line with psychometric work on SEL surveys conducted by Meyer et al. (2018) and Davidson et al. (2016), we use these terms since they are accepted terms for this type of analysis.<sup>19</sup> In the case of the SEC and CC items, these terms are used to indicate whether most respondents indicated less favorable or more favorable perceptions of their SEC or the school CC based on Likert-type response options. The item difficulty metric provides useful information in that it indicates whether the item contributes meaningful information about the relevant construct being measured.

Table 2 provides the list of items that are identified for further analysis based on the high or low proportion of students in our analytic sample that endorsed each survey item. Looking at endorsement rates by grade level, we do not find any item with endorsements below the threshold of 0.30, but we find 14 items with endorsements above the common threshold of 0.80, indicating that most students have favorable perceptions of the factor being assessed. Table 2 also shows that the majority of items with above 80 percent endorsement belong to one SEC scale (i.e., self-regulation) and two CC scales (i.e., rigorous expectations and school safety).

---

<sup>19</sup> See Washoe County School District’s Social Emotional Learning webpage for a practical application of this type of analysis: <http://www.wcsddata.net/data-topics/sel/>

While we flag the items below for further diagnostics, we consider multiple aspects of item characteristics to decide whether any of these items are potentially problematic to include in future surveys.

**Table 2.** Item Response Spread: Items identified for further diagnostics

Construct	Survey	Item Wording	Spread Range	Grades Flagged
Growth Mindset	SEC	If I am not naturally smart in a subject, I will never do well in it.*	0.84	4
Self-Regulation	SEC	I came to class prepared.	0.81 - 0.84	5, 11, 12
Self-Regulation	SEC	I remembered and followed directions.	0.83 - 0.85	11, 12
Self-Regulation	SEC	I allowed others to speak without interruption.	0.82	12
Self-Regulation	SEC	I was polite to adults and peers.	0.81 - 0.87	4, 5, 6, 11, 12
Social Awareness	SEC	I cared about other people's feelings.	0.82	4
Rigorous Expectations	CC	When you feel like giving up on a difficult task, how likely is it that your teachers will make you keep trying?	0.83	4
Rigorous Expectations	CC	How much do your teachers encourage you to do your best?	0.80 - 0.84	4, 5
Rigorous Expectations	CC	How often do your teachers take time to make sure you understand the material?	0.81	4
School Safety	CC	How often do students get into physical fights at your school?*	0.82 - 0.87	9, 10, 12
School Safety	CC	How likely is it that someone from your school will bully you online?*	0.81 - 0.90	4, 5, 8-12
School Safety	CC	How often do you worry about violence at your school?*	0.81 - 0.87	9, 10, 11, 12
School Safety	CC	If a student is bullied in school, how difficult is it for him/her to get help from an adult?*	0.82	12
Teacher-Student Relationship	CC	How many of your teachers are respectful towards you?	0.80 - 0.83	4, 5, 11

\*Reverse-coded item

### Item discrimination

We examine item discrimination using item-total correlation based on classical test theory. Item-total correlation is the correlation between a given item and all the items in the scale. It is expected that a student who responds to an item with a 4 or 5 on a Likert scale should also have, in general, higher overall scores on the construct compared to a student who responds with a 2 or 3 on that item. This metric indicates how well an item is able to discriminate between students with low and high perceptions of the construct being measured.

We find that all items have adequate discrimination, with item-total correlations much greater than the common threshold of 0.40 for survey responses (Ladhari, 2010). On the SEC survey, item-total correlations range from 0.64 to 1 among students in grades 4-12. For the CC survey, item-total correlations range from 0.52 to 0.97 for students in grades 4-12.

### **Omit rate**

Examining omit rates for each survey item at each grade level allows us to identify items that are skipped by a large proportion of students. In cases where students are systematically omitting an item or entire scale, it can bias our interpretation of the aggregate response of a scale. We find high response rates at the item level across both SEC and CC surveys, with at least 96 percent of students responding to all items in the survey. For the four percent of data that is missing, we do not find any consistent patterns in which items are skipped by grade level. We find that students in grade 4 and grade 6 tend to skip more items throughout the survey than any other grade. In grade 4, the self-regulation item “I remained calm even when criticized or otherwise provoked” (Q6) is the most commonly skipped item. In grade 6, the curiosity items “I like activities that challenge my thinking abilities” (Q2) and “I find satisfaction in thinking hard and for long hours” (Q4) are the most commonly skipped items. Other items with relatively high skip rates among sixth graders include “Overall, how well do your learning strategies help you learn more effectively?” (learning strategies Q4); “My intelligence is something that I can’t change very much” (growth mindset Q3); and “I am diligent. I never give up” (perseverance Q4). Nonetheless, responses rates for these item/grade combinations are still quite high, ranging

from 96 to 100 percent.<sup>20</sup> One hypothesis is that the items skipped most frequently are those that contain words or concepts that may be difficult for students to understand.

### **Differential Item Functioning (DIF)**

We conduct DIF analysis using the standardized mean difference (SMD) method in conjunction with the Mantel chi-square statistic (Dorans & Kulick, 1986; Dorans & Schmitt, 1991; Mantel, 1963; Mantel & Haenszel, 1959). The SMD method was originally developed by Dorans and Kulick (1986) to estimate DIF for dichotomous items. Dorans and Schmitt (1991) extended this statistic to estimate DIF for items with more than two possible scores (e.g., Likert-type response choices like those used in our surveys). An item is classified as having moderate-to-large DIF if the Mantel chi-square *p*-value is below 0.05 and the ratio of absolute value of SMD and the standard deviation of the item is higher than 0.25 (Dorans & Schmitt, 1991; Zwick & Ercikan, 1989). To ensure a large enough *n*-size, we combine adjacent grade levels: 4-5 (elementary), 6-8 (middle), 9-10 (early high school), and 11-12 (late high school), and calculate DIF statistics only for those where both the focal group and the reference group have at least 100 students who completed all items within the scale. Table 3 lists the focal and standard reference comparison groups for DIF.<sup>21</sup>

---

<sup>20</sup> We do not look at omit rates for demographic subgroups because of high response rates overall and at the grade level.

<sup>21</sup> The reference group is typically chosen based on the types of differential item functioning one is trying to disprove (Holland and Wainer, 1993). Historically, the normativistic models based on White, middle-class children have been dominant in the fields of developmental psychology and child development which can lead to survey development which privileges the experience of White male students above others. Following from this, our primary concern is that the items were designed based on a White- and male-dominant conception of strong SEC and favorable school CC. To test whether the survey is biased toward these groups, we include White students as our reference group when testing DIF by race/ethnicity, and we include males as our reference group when testing DIF by gender.

**Table 3. DIF Comparison Groups**

<b>Grouping Variable</b>	<b>Focal Group</b>	<b>Reference Group</b>
Gender	Female	Male
Ethnicity	Black or African American Latino	White White
Socioeconomic Status	Students who receive free/reduced-price lunch	Students who do not receive free/reduced-price lunch
Special Instructional Needs	English language learners (ELL) Students with disabilities	Non-ELL Students without disabilities
Administration Term	Fall	Spring

An item with moderate-to-high DIF is often considered a concern when it is measuring different abilities for the focal group after controlling for overall differences in survey scores (Dorans & Schmitt, 1991). Table 4 provides a summary of the DIF items and the respective focal subgroups where we find moderate-to-high DIF from survey administrations in the Fall of 2016 and/or Spring of 2017. Of the 39 SEC items, we find four items with moderate-to-high DIF for one or more focal group(s) based on the threshold specified above. For example, one self-regulation item, “I kept my temper in check,” was identified in both the Fall and Spring but for different focal groups and at different grade levels. Similarly, one social awareness item, “I cared about other people’s feelings,” was identified as having moderate-to-high DIF in Spring 2017 for Black/African American students in middle school.

Among the 36 culture-climate items, five of the six school safety items were identified in different DIF analyses across elementary and middle school grade levels and among different subgroup comparisons. Of those five school safety items, the item “How often do you worry about violence at your school?” was identified in three different demographic subgroup analyses with moderate-to-high DIF for ELL, FRPL, and Latino students. Similarly, the item “If a student is bullied in school, how difficult is it for him/her to get help from an adult?” in the school safety scale had moderate-to-high DIF for ELL students in elementary school, middle school, and early

high school in the Fall. It also had moderate-to-high DIF for ELL students in middle school in the Spring and for girls in early high school in the Spring.

**Table 4.** Items with moderate-to-high DIF that are differentially difficult for focal group(s)

Survey	Construct	Item	Moderate-to-high DIF for Subgroup	Grade Level	Term
SEC	Curiosity	I like complex problems more than easy problems.	FRPL	Elementary	Fall 2016
	Curiosity	I enjoy situations where I will have to think about something.	SPED	Early HS	Fall 2016
	Growth Mindset	My intelligence is something that I can't change very much.	FRPL; Black or African American	Middle School	Fall 2016
	Self-Awareness	Knowing the emotions I feel.	Female	Late HS	Spring 2017
	Self-Efficacy	I can earn an A in my classes.	FRPL	Early HS	Spring 2017
	Self-Regulation	I was polite to adults and peers.	Black or African American	Elementary	Fall 2016
	Self-Regulation	I kept my temper in check.	Female; Black or African American	Middle; Late HS	Both
	Social Awareness	I cared about other people's feelings.	Black or African American	Middle School	Spring 2017
	Social Awareness	I was able to describe my thoughts and feelings in ways that others understood.	Female	Early HS	Fall 2016
Social Awareness	I respected the views of others even if I disagreed with them.	Black or African American	Elementary	Fall 2016	
CC	Engagement	In your classes, how eager are you to participate?	Latino	Middle School	Fall 2016
	Rigorous Expectations	How often do your teachers make you explain your answers?	ELL	Middle; Early HS	Fall 2016
	School Safety	How often do students get into physical fights at your school?	Latino	Middle School	Spring 2017
	School Safety	How likely is it that someone from your school will bully you online?	Female	Late HS	Fall 2016
	School Safety	How often do you worry about violence at your school?	ELL; FRPL; Latino; Black or African American	Elementary; Middle	Both
	School Safety	If a student is bullied in school, how difficult is it for him/her to get help from an adult?	ELL; Female	Elementary; Middle; Early HS	Both
	School Safety	At your school, how unfairly do the adults treat the students?	Black or African American	Elementary	Fall 2016
	Teacher-Student Relationship	If you came back to visit class three years from now, how many of your teachers would be excited to see you?	Latino	Middle School	Spring 2017

## Factor Analysis

We use exploratory factor analysis (EFA) to examine the dimensionality of the full surveys measuring SEC and CC. EFA allows us to uncover the internal structure of a survey by indicating how groups of items hang together within the full instrument. We then proceed with confirmatory factor analysis (CFA) to validate the hypothesized model structure and examine “goodness of fit” indices, including the comparative fit index (CFI), Tucker-Lewis Index (TLI), root-mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR).

The first step in factor analysis is to determine whether the data is suitable for factor analysis. We use the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1974) and Bartlett’s test of sphericity to assess the suitability of data. The KMO is a statistic that allows us to measure the proportion of variance among our survey items that might be indicative of underlying latent common constructs. The KMO test returns values between 0 and 1, with values greater than 0.80 indicating adequate sampling. Bartlett’s test approximates a chi-square distribution and tests the hypothesis that the variables being analyzed are independent. Data is deemed appropriate for factor analysis if the  $p$ -value is less than 0.05. We find that our survey data satisfies both KMO (0.96) and Bartlett’s test ( $p < 0.05$ ) criteria, suggesting that our sample is adequate and appropriate for factor analysis.

**Exploratory factor analysis.** After testing for sampling adequacy and data suitability, we conduct EFA to empirically examine the structure of the survey questions. EFA allows us to explore the factor structure of SEC and CC constructs that emerge from the data without *a priori* specifying any theories about the constructs. It is a common rule-of-thumb in EFA to use factor loadings of 0.30 or greater to identify practically significant factor loading (Schmitt & Sass,

2011). We use this criteria to identify factor structures and loadings of items to factors for both SEC and CC.

Table 5 presents the factor structure and loadings of items to factors for the SEC survey. As it illustrates, our EFA examination results in seven factors. Four of the seven factors clearly group into distinct scales forming the growth mindset, self-awareness, self-efficacy, and social awareness scales. However, three of the factors—self-regulation, perseverance, and curiosity—have overlapping items loading onto two factors. Specifically, three perseverance items (Q1, Q2, Q3) load onto self-regulation and two perseverance items (Q3, Q4) load onto curiosity.

Table 6 presents the factor structure and loadings of items to factors for the CC survey. As seen in Table 6, three of the seven factors clearly group into three scales, forming the learning strategies, engagement, and rigorous expectations scales. However, four of the factors, namely cultural and linguistic competence, teacher-student relationship, school safety, and sense of belonging, have overlapping items loading onto two factors. Specifically, one school safety item loads slightly more strongly (0.32) onto cultural and linguistic competence compared to school safety (0.29) and one sense of belonging item loads more strongly (0.38) onto teacher-student relationship compared to sense of belonging (0.23).

**Confirmatory factor analysis.** We use CFA to test the *a priori* hypothesized seven-factor model based on seven SEC and seven CC scales that were curated from the prior surveys. Using Spring 2017 survey data, we fit two models: one for the SEC constructs and the other for the CC constructs. We use an R package called *lavaan* to run CFA models at each grade level separately. We treat the data as ordinal, use a three-stage weighted least squares (WLS) estimation approach, and set the constructs to be correlated (Forero et al., 2009).

We use the following criteria to evaluate model fit (Holgado-Tello et al., 2010): (1) the comparative fit index (CFI) to control for the effects of sample size; (2) the root mean square error of approximation (RMSEA) to examine how well the model fits the populations covariance matrix; (3) the Tucker-Lewis Index to assess the correspondence between the proposed model and the data; and (4) the standardized root mean square residual (SRMR). Acceptable threshold for these indices are as follows: 0.95 or larger for CFI and TLI; 0.06 or smaller for RMSEA; and 0.08 or smaller for SRMR (Hu & Bentler, 1999; Schreiber et al., 2006).

Table 7 provides details about fit indices for both surveys at each grade level. Results of CFA indicate that the hypothesized model appears to be a good fit to the data. Grade-level analyses of the four fit indices for grades 4 to 12 confirm good fit, with overall indices performing within acceptable thresholds. For the SEC and CC models, the CFI indices range from 0.98 to 1.00; TLI ranges from 0.98 to 1; RMSEA ranges from 0 to 0.06; and SRMR ranges from 0.05 to 0.08. Our examination of modification index shows that all SEC items have modification indices below 100, which is the acceptable threshold (Hooper et al., 2008). At the grade level, a few items have modification indices above 100 in some grade levels, but there is no consistent pattern across grades.

**Table 5.** Exploratory Factor Analysis of SEC Student Survey: Factor Loadings with Oblimin Rotation

Item Wording	Factor Loadings						
	1	2	3	4	5	6	7
<b>Self-Regulation</b>							
I came to class prepared.	0.63						
I remembered and followed directions.	0.72						
I got my work done right away instead of waiting until the last minute.	0.60						
I paid attention and resisted distractions.	0.67						
I worked independently with focus.	0.66						
I remained calm even when criticized or otherwise provoked.	0.37						0.36
I allowed others to speak without interruption.	0.50						
I was polite to adults and peers.	0.51						
I kept my temper in check.	0.30						0.47
<b>Self-Efficacy</b>							
I can earn an A in my classes.						0.75	
I can do well on all my tests, even when they are difficult.						0.90	
I can master the hardest topics in my classes.						0.85	
I can meet all the learning goals my teachers set.						0.68	
<b>Perseverance</b>							
I finish whatever I begin.	0.45	0.11					
I work very hard. I keep working when others stop to take a break.	0.45	0.28					
I stay interested in my goals, even if they take a long time (months or years) to complete.	0.32	0.39					
I am diligent. I never give up.	0.27	0.38					
<b>Growth Mindset</b>							
Challenging myself won't make me any smarter.			0.73				
There are some things I am not capable of learning.			0.78				
My intelligence is something that I can't change very much.			0.78				
If I am not naturally smart in a subject, I will never do well in it.			0.75				
<b>Curiosity</b>							
I like complex problems more than easy problems.		0.71					
I like activities that challenge my thinking abilities.		0.77					
I enjoy situations where I will have to think about something.		0.78					
I find satisfaction in thinking hard and for long hours.		0.81					
I enjoy thinking about new solutions to problems.		0.75					
I like to think of my life as a puzzle that I must solve.		0.61					
<b>Self Awareness</b>							
Knowing what my strengths are.					0.57		
Knowing ways I calm myself down.					0.63		
Knowing when my feelings are making it hard for me to focus.					0.71		
Knowing the emotions I feel.					0.76		
<b>Social Awareness</b>							
I listened carefully to other people's points of view.				0.57			
I cared about other people's feelings.				0.82			
I noticed and complimented others' accomplishments.				0.81			
I got along with students who were different from me.				0.70			
I was able to describe my thoughts and feelings in ways that others understood.				0.55			
I respected the views of others even if I disagreed with them.				0.73			
I was able to stand up for myself without putting others down.				0.57			
I knew how to disagree without starting an argument.				0.45			

**Table 6.** Exploratory Factor Analysis of CC Student Survey: Factor Loadings with Oblimin Rotation

Item Wording	Factor Loadings						
	1	2	3	4	5	6	7
<b>Cultural and Linguistic Competence</b>							
This school provides instructional materials that reflect my cultural background.						0.41	
Adults working at this school treat all students respectfully.						0.60	
People of different cultural backgrounds, races or ethnicities get along well at						0.63	
All students are treated the same, regardless of whether their parents are rich or poor.						0.75	
Boys and girls are treated equally well.						0.76	
<b>Teacher-Student Relationship</b>							
How many of your teachers are respectful towards you?	0.47						
If you walked into class upset, how many of your teachers would be concerned?	0.66						
If you came back to visit class three years from now, how many of your teachers would be excited to see you?	0.78						
When your teachers ask how you are doing, how many of them are really interested in your answer?	0.73						
How many of your teachers would you be excited to have again in the future?	0.71						
<b>School Safety</b>							
How often are people disrespectful to others at your school?		0.48					
How often do students get into physical fights at your school?		0.67					
How likely is it that someone from your school will bully you online?		0.50					
How often do you worry about violence at your school?		0.77					
If a student is bullied in school, how difficult is it for him/her to get help from an adult?		0.52					
At your school, how unfairly do the adults treat the students?		0.29				0.32	
<b>Sense of Belonging</b>							
How well do people at your school understand you as a person?				0.59			
How connected do you feel to the adults at your school?	0.38			0.23			
How much respect do students in your school show you?				0.60			
How much do you matter to others at this school?				0.64			
Overall, how much do you feel like you belong at your school?				0.41			
<b>Learning Strategies</b>							
When you get stuck while learning something new, how likely are you to try a different strategy?			0.66				
How confident are you that you can choose an effective strategy to get your schoolwork done well?			0.73				
Before you start on a challenging project, how often do you think about the best			0.66				
Overall, how well do your learning strategies help you learn more effectively?			0.71				
How often do you use strategies to learn more effectively?			0.81				
<b>Engagement</b>							
How excited are you about going to your classes?					0.80		
How often do you get so focused on activities in your classes that you lose track of time?					0.59		
In your classes, how eager are you to participate?					0.65		
When you are not in school, how often do you talk about ideas from your classes?					0.65		
Overall, how interested are you in your classes?					0.83		
<b>Factor 7: Rigorous Expectations</b>							
How often do your teachers make you explain your answers?							0.64
When you feel like giving up on a difficult task, how likely is it that your teachers will make you keep trying?							0.67
How much do your teachers encourage you to do your best?							0.72
How often do your teachers take time to make sure you understand the material?							0.59
Overall, how high are your teachers' expectations of you?							0.59

**Table 7.** Confirmatory Factor Analysis Model Fit Indices by Grade

	Student-Reported SEC (Spring 2017)				Student-Reported CC (Spring 2017)			
	CFI	TLI	RMSEA	SRMR	CFI	TLI	RMSEA	SRMR
Grade 4	0.995	0.995	0.030	0.062	0.985	0.983	0.042	0.068
Grade 5	0.993	0.992	0.042	0.057	0.989	0.988	0.044	0.057
Grade 6	0.997	0.996	0.040	0.051	0.989	0.988	0.060	0.063
Grade 7	0.992	0.991	0.051	0.055	0.989	0.988	0.056	0.057
Grade 8	0.993	0.992	0.048	0.066	0.991	0.991	0.045	0.070
Grade 9	0.991	0.990	0.056	0.056	0.990	0.989	0.053	0.057
Grade 10	0.986	0.985	0.067	0.068	0.990	0.989	0.052	0.064
Grade 11	0.982	0.980	0.059	0.088	0.989	0.987	0.043	0.081
Grade 12	1.000	1.000	0.000	0.079	1.000	1.000	0.000	0.082

### Discussion

In this study, we examined the measurement properties of the SEC and CC surveys administered to students in the NewSchools Invent cohort of schools, using data from the Fall of 2016 and Spring of 2017. We find that both SEC and CC scales demonstrate high internal consistency and that the items appear to discriminate appropriately across students’ perceptions of their SEC and school CC. For a few select items, most students appear to be answering them in the same way, suggesting that the items do not provide differentiating information. However, since there is no consistent pattern across grade levels, and the items fit well to expected factor structures, we will continue to include these items in the survey and re-examine them after we have additional data from subsequent years of the partnership. If, at that time, we continue to see evidence that the identified items do not provide differentiating information, we will explore ways to improve the variability in student responses without changing the underlying concept the item is intended to measure.

We have identified select items that show evidence of differential item functioning across subgroups and between administrations; however, we do not believe immediate change to the

items is warranted. First, there is a lack of consistency among most items identified for DIF across subgroups and grade levels. We need more data and more evidence in order to draw conclusions on how and why these items function differently between student subgroups (Scott et al., 2009). Therefore, we will re-examine items when we have additional years of data, and flag those identified for DIF consistently across multiple terms, grade levels and years. Second, DIF is intended to highlight items on which students from different subgroups have differential responses, holding constant their overall scores on the scale or survey. However, an item flagged for DIF does not necessarily imply item bias. For example, differential functioning of the CC items may highlight important differences across subgroups in students' perceptions or experiences of a particular aspect of the school culture or climate (Raju et al., 2002). To understand whether differential functioning suggests bias, additional evidence is required; for example, evidence that an item contains content that is irrelevant to the construct being measured or that the item assesses content relevant only to a particular cultural conception of that construct (Zumbo, 1999). We intend to investigate the selected items in greater detail through future analyses with additional years of data and through student cognitive interviews, which can help illuminate *why* students from different subgroups are responding differently on particular items.

Exploratory factor analysis indicates potential cross-loadings of some items from the perseverance scale onto the self-regulation scale, suggesting there may be overlap in the underlying constructs being assessed. In fact, both constructs are viewed by personality psychologists as sub-facets of conscientious (Poropat, 2009). Further, Wolters & Hussain (2014) found that perseverance of efforts under the Grit Short Scale (Duckworth & Quinn, 2009) was consistently associated with all indicators of the self-regulated learning scale. This association

points to the possibility that the perseverance and self-regulation items may be conceptually similar.

We also find cross-loading of two items from the perseverance scale onto the curiosity scale, suggesting that items from each scale are measuring the same underlying construct. This could be due to the fact that the identified items from the perseverance scale are quite similar to those in the curiosity scale. For example, an item from the perseverance scale, “I work very hard. I keep working when others stop to take a break,” is quite similar conceptually to an item from the curiosity scale, “I find satisfaction in thinking hard and for long hours.” Alternatively, it may be that students are responding similarly to both scales because all curiosity and perseverance items are grouped together in the same section on the survey. To explore the latter possibility, we have changed the survey for one-third of the students in the Spring 2018 test administration, in which the perseverance items and curiosity items have been placed into separate survey sections. We will then examine factor structure for those respondents for whom the survey remained the same, and compare it to the factor structure for those respondents for whom the two scales were placed in distinct sections of the survey.

Overall, results from the EFA and CFA suggest that our survey scales are capturing distinct constructs. Researchers recommend making modifications only when it makes theoretical sense and other fit indices also do not perform as expected (Schreiber et al., 2006). Because of the fit of the data to the model, we recommend keeping the design of the survey as is and continuing to investigate the fit when additional years of data become available. We further advise against immediate changes to the included scales since the survey is intended to serve as a common measure across schools based on prioritized constructs identified by school leaders. As

such, additional input from these school leaders would be warranted before removing or modifying a survey scale.

### **Conclusion**

While the individual SEC and CC scales were curated from existing surveys, this is the first study to examine the measurement properties of these scales when administered as cohesive surveys of student SEC and school CC in a unique sample of schools that are part of the NewSchools Invent cohort. Our preliminary results suggest that the SEC and CC surveys scales are suitable for practitioners to use to inform specific classroom strategies or instructional practices. Of course, validation is an ongoing process in which multiple sources of evidence should be brought to bear (Kane, 2006; 2013; Messick, 1989, 1995). In this paper, we focused primarily on examining how well the surveys were designed – that is, the extent to which the items provide consistent and new information about the underlying constructs being assessed, the extent to which the items are interpreted comparably across student subgroups, and the extent to which scales for each survey are measuring unique constructs. Based on our results, there are two primary aspects of survey design that require further investigation: whether particular items are truly interpreted differentially across subgroups (and if so, why), and the extent to which select SEC scales are measuring overlapping constructs. We will continue to explore both questions using qualitative and quantitative analyses of additional years of data. Finally, through an ambitious research and learning agenda with our NewSchools partnership, we will continue to accumulate validity evidence in order to better understand survey strengths and limitations for practitioner use when making data-informed decisions within and across schools.

## References

- Anderman, L. H. (2003). Academic and social perceptions as predictors of change in middle school students' sense of school belonging. *The Journal of Experimental Education, 72*(1), 5-22.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation, 19*(5), 2.  
<http://pareonline.net/pdf/v19n5.pdf>
- Berg, J., Osher, D., Same, M. R., Nolan, E., Benson, D., & Jacobs, N. (2017). *Identifying, defining, and measuring social and emotional competencies*. Washington, DC: American Institutes for Research.
- Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*(3), 899-911.
- Buckley, K. & Krachman, S. (2016). Patterns in student self-report and teacher report measures of social emotional mindsets, skills and habits: Initial findings from the Boston charter research collaborative. Boston: Transforming Education. Retrieved from  
<https://www.transformingeducation.org/wp-content/uploads/2017/04/TE-BCRCWorkingPaperFINAL.pdf>
- Claudia, P., & Raju, N. S. (2002). A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory. Presented at the National Council on Measurement in Education (NCME) Annual Meeting, New Orleans, Louisiana.
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin, 103*(2), 276.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.
- Davidson, L. A., Crowder, M. K., Gordon, R. A., Domitrovich, C. E., Brown, R. D., & Hayes, B. I. (2016). A continuous improvement approach to social and emotional competency measurement. *Journal of Applied Developmental Psychology, 55*, 93-106.
- Dorans, N. J., & Schmitt, A. P. (1991). Constructed response and differential item functioning: A pragmatic approach. *ETS Research Report Series, 1991*(2).

- Duckworth, A. L. (2008). Self-control as a multi-dimensional construct: Results from a meta-analysis. *International Journal of Psychology, 43*(3), 12.
- Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science, 11*(1), 35-55.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237-251.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of non-cognitive factors in shaping school performance--a critical literature review*. Chicago, IL: Consortium on Chicago School Research.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*(4), 625-641.
- Gabrieli, C., Ansel, D., & Krachman, S. B. (2015). Ready to be counted: The research case for education policy action on non-cognitive skills. Boston: Transforming Education. Retrieved from [https://www.transformingeducation.org/wp-content/uploads/2017/04/ReadytoBeCounted\\_Release.pdf](https://www.transformingeducation.org/wp-content/uploads/2017/04/ReadytoBeCounted_Release.pdf)
- Gehlbach, H. & Hough, H. (2018). Measuring social emotional learning through student surveys in the core districts: a pragmatic approach to validity and reliability. Palo Alto, CA: Policy Analysis for California Education. Retrieved from [http://edpolicyinca.org/sites/default/files/SEL\\_Validity\\_May-2018.pdf](http://edpolicyinca.org/sites/default/files/SEL_Validity_May-2018.pdf).
- Guerra, N. G., & Jagers, R. (1998). The importance of culture in the assessment of children and youth. In *Studying minority adolescents: Conceptual, methodological, and theoretical issues*, 167-181. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.
- Hanson, T., & Kim, J. (2007). Measuring resilience and youth development: The psychometric properties of the Healthy Kids Survey. *Issues & Answers. REL 2007-No. 34*. San

- Francisco: Regional Educational Laboratory West. Retrieved from <https://eric.ed.gov/?id=ED498459>
- He, Q., & Wheadon, C. (2013). The effect of sample size on item parameter estimation for the partial credit model. *International Journal of Quantitative Research in Education*, 1(3), 297-315. See page 14 of this paper for discussion of sample size criteria.
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903-918.
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153-166.
- Holland, P. W. & Wainer, H. (1993). *Differential Item Functioning*. New York, NY: Routledge Taylor & Francis Group.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 2. Retrieved from [www.ejbrm.com](http://www.ejbrm.com)
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Jackson, J. J., Connolly, J. J., Garrison, S. M., Leveille, M. M., & Connolly, S. L. (2015). Your friends know how long you will live: A 75-year study of peer-rated personality traits. *Psychological Science*, 26(3), 335-340.
- Jeong, H. J., & Lee, W. C. (2016). Does differential item functioning occur across respondents' characteristics in safety attitudes questionnaire. *Biometrics & Biostatistics International Journal*, 4(3), 00097.
- Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, 7, 109.

- Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health, 105*(11), 2283-2290.
- Jones, S. M., & Kahn, J. (2017). The evidence base for how we learn: Supporting students' social, emotional, and academic development. Washington, DC: Aspen Institute.  
Retrieved from <https://eric.ed.gov/?id=ED577039>
- Kaiser, M. O. (1974). Kaiser-Meyer-Olkin measure for identity correlation matrix. *Journal of the Royal Statistical Society, 52*, 296-298.
- Kane, M. (2006). Content-related validity evidence in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (131-153). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Educational Measurement, 50*, 1-73.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Konishi, C., Hymel, S., Zumbo, B. D., & Li, Z. (2010). Do school bullying and student—teacher relationships matter for academic achievement? A multilevel analysis. *Canadian Journal of School Psychology, 25*(1), 19-39.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537-567.
- Ladhari, R. (2010). Developing e-service quality scales: A literature review. *Journal of Retailing and Consumer Services, 17*(6), 464-477.
- Lamb, R. L., Annetta, L., Meldrum, J., & Vallett, D. (2012). Measuring science interest: Rasch validation of the science interest survey. *International Journal of Science and Mathematics Education, 10*(3), 643-668.
- LaRocca, R. & Krachman, S.B. (2017). Expanding the Definition of student Success under ESSA: Opportunities to Advance Social-Emotional Mindsets, Skills, and Habits for Today's Students, <https://www.transformingeducation.org/wp-content/uploads/2017/05/Expanding-Success-Policy-Brief-edit.pdf>
- Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58*(303), 690-700.

- McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471-490.
- McColskey, F. J., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring student engagement in upper elementary through high school: A description of 21 instruments. *Issues & Answers. REL 2011-No. 098*. San Francisco: Regional Educational Laboratory Southeast. Retrieved from [https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL\\_2011098.pdf](https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2011098.pdf)
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-104). New York, NY: American Council on Education and Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Meyer, R., Wang, C., & Rice, A. (2018). Measuring students' social-emotional learning among California's CORE districts: An IRT modelling approach. Stanford, CA: Policy Analysis for California Education (PACE). Retrieved from <http://www.edpolicyinca.org/publications/sel-measurement>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Sears, M. R. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693-2698.
- Pagano, I. S., & Gotay, C. C. (2005). Ethnic differential item functioning in the assessment of quality of life in cancer patients. *Health and Quality of Life Outcomes, 3*(1), 60.
- Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1-18.
- Pearson. (2017). *PARCC: Final technical report for 2016 administration*. Retrieved from <https://parcc-assessment.org/research/>
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

- Phillips, A. W., Reddy, S., & Durning, S. J. (2016). Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. *Medical Teacher*, 38(3), 217-228.
- Phillips, S. F., & Rowley, J. F. (2015). The Tripod School Climate Index: An invariant measure of school safety and relationships. *Social Work Research*, 40(1), 31-39.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin* 135(2), 322–338
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517.
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., & Mazur, J. (2008). The KIDSCREEN-52 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Value in Health*, 11(4), 645-658.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4(3), 207-230.
- Rogers, H. J. (2014). Differential item functioning. *Wiley StatsRef: Statistics Reference Online*, 1-7. Retrieved from <https://onlinelibrary.wiley.com/doi/full/10.1002/9781118445112.stat06381.pub2>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338.
- Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, 71(1), 95-113.
- Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., ... & Sprangers, M. A. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62(3), 288-295.

- Squires, J., Estabrooks, C., O'Rourke, H., Gustavsson, P., Newburn-Cook, C., & Wallin, L. (2011). A systematic review of the psychometric properties of self-report research utilization measures used in healthcare. *Implement Science*, 6(83).
- Starkweather, J. (2012). Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability. *Benchmarks RSS Matters*. Retrieved from [https://it.unt.edu/sites/default/files/omega\\_jds\\_jun2012.pdf](https://it.unt.edu/sites/default/files/omega_jds_jun2012.pdf)
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53.
- Transforming Education (2016). Measuring MESH, student and teacher surveys curated for the CORE districts. Retrieved from <https://www.transformingeducation.org/measuring-mesh/>.
- West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2017). Development and implementation of student social-emotional surveys in the CORE Districts. *Journal of Applied Developmental Psychology*, 55, 119-129.
- Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.8594&rep=rep1&type=pdf>
- Woods, C. M., & Grimm, K. J. (2011). Testing for non-uniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339-361.
- Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). Executive function: Implications for education. NCER 2017-2000. *National Center for Education Research*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from <https://ies.ed.gov/ncer/pubs/20172000/pdf/20172000.pdf>
- Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). Ottawa, ON: National Defense Headquarters.
- Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 26(1), 55-66.

## Appendix A

**Table A1. Survey Items and Scales: Student-reported Social Emotional Competencies**

Scale	Items	Response Options	Source
Curiosity	I like complex problems more than easy problems.	Not at all like me to Very much like me	NAEP – General Student Questionnaire 2016 Pilot  Link to source (p. 15): <a href="https://nces.ed.gov/nationsreportcard/subject/about/pdf/bgq/student/2016_bq_student_g4_core_p.pdf">https://nces.ed.gov/nationsreportcard/subject/about/pdf/bgq/student/2016_bq_student_g4_core_p.pdf</a>
	I like activities that challenge my thinking abilities.		
	I enjoy situations where I will have to think about something.		
	I find satisfaction in thinking hard and for long hours.		
	I enjoy thinking about new solutions to problems.		
	I like to think of my life as a puzzle that I must solve.		
Growth Mindset <i>(reverse coded)</i>	Challenging myself won't make me any smarter.	Not at all true to Completely true	CORE & BCRC -- Farrington et al. (2013) Becoming Effective Learners Survey Development Project, Chicago Consortium for School Research.  Link to paper: <a href="https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive%20Report.pdf">https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive%20Report.pdf</a>
	There are some things I am not capable of learning.		
	My intelligence is something that I can't change very much.		
	If I am not naturally smart in a subject, I will never do well in it.		
Perseverance	I finish whatever I begin.	Not at all like me to Very much like me	BCRC – Adapted from Duckworth  Link to paper: <a href="https://globaled.gse.harvard.edu/files/geii/files/validation_grit_scale_duckworth_jpa_m_figueroa-2.pdf">https://globaled.gse.harvard.edu/files/geii/files/validation_grit_scale_duckworth_jpa_m_figueroa-2.pdf</a>
	I work very hard. I keep working when others stop to take a break.		
	I stay interested in my goals, even if they take a long time (months or years) to complete.		
	I am diligent. I never give up.		

			<p>Link to source:  <a href="https://www.dropbox.com/s/rn5wo3y0iis0qtf/8-item%20Grit%20Scale_Child%20Adapted%20Version_4.pdf?dl=0">https://www.dropbox.com/s/rn5wo3y0iis0qtf/8-item%20Grit%20Scale_Child%20Adapted%20Version_4.pdf?dl=0</a></p>
Self-Awareness	Knowing what my strengths are.	Very difficult to Very easy	<p>Washoe – 2016 WCSO-CASEL Long Version Items, developed through IES grant</p> <p>Link to source:  <a href="https://www.washoeschools.net/Page/10932">https://www.washoeschools.net/Page/10932</a></p>
	Knowing ways I calm myself down.		
	Knowing when my feelings are making it hard for me to focus.		
	Knowing the emotions I feel.		
Self-Efficacy	I can earn an A in my classes.	Not at all confident to Completely confident	<p>CORE &amp; BCRC -- Adapted from Farrington et al. (2014) <i>Becoming Effective Learners Survey Development Project, Chicago Consortium for School Research.</i></p> <p>Link to paper:  <a href="https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive%20Report.pdf">https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive%20Report.pdf</a></p>
	I can do well on all my tests, even when they are difficult.		
	I can master the hardest topics in my classes.		
	I can meet all the learning goals my teachers set.		
Self-Regulation	I came to class prepared.	Almost never to Almost all the time	<p>CORE &amp; BCRC -- Adapted from Patrick &amp; Duckworth (2013, May) <i>Empirical support for a tripartite taxonomy of character in adolescents.</i> Poster presented at the 25th annual convention of the Association for Psychological Science.</p>
	I remembered and followed directions.		
	I got my work done right away instead of waiting until the last minute.		
	I paid attention and resisted distractions.		
	I worked independently with focus.		
	I remained calm even when criticized or otherwise provoked.		

	<p>I allowed others to speak without interruption.</p> <p>I was polite to adults and peers.</p>		<p>Link to paper: <a href="https://www.researchgate.net/publication/280529814_A_tripartite_taxonomy_of_character">https://www.researchgate.net/publication/280529814_A_tripartite_taxonomy_of_character</a></p> <p>Link to source: <a href="https://www.dropbox.com/s/stll4oeia61mm9k/DSIS-C.pdf?dl=0">https://www.dropbox.com/s/stll4oeia61mm9k/DSIS-C.pdf?dl=0</a></p>
	<p>I kept my temper in check.</p>		
Social Awareness	<p>I listened carefully to other people's points of view.</p> <p>I cared about other people's feelings.</p> <p>I noticed and complimented others' accomplishments.</p> <p>I got along with students who were different from me.</p> <p>I was able to describe my thoughts and feelings in ways that others understood.</p> <p>I respected the views of others even if I disagreed with them.</p> <p>I was able to stand up for myself without putting others down.</p> <p>I knew how to disagree without starting an argument.</p>	<p>Almost never to Almost all the time</p>	<p>BCRC – Adapted from CASEL &amp; AIR / Washoe County Items in 2013</p> <p>Link to source: <a href="https://www.washoeschools.net/Page/10932">https://www.washoeschools.net/Page/10932</a></p>

**Table A2. Survey Items and Scales: Student-Reported Culture-Climate**

Scale	Items	Response Options	Source
Cultural and Linguistic Competence <i>(reverse coded)</i>	This school provides instructional materials that reflect my cultural background, ethnicity and identity.	Strongly Agree to Strongly Disagree	The cultural and linguistic competence scale was drawn from the EDSCLS. It is based on a 4-point Likert Scale. For the purposes of our survey, we converted it to a 5-point Likert scale to be consistent with the other six CC scales drawn from Panorama.  Link to source: <a href="https://safesupportivelearning.ed.gov/sites/default/files/EDSCLS_Questionnaires_112_017.pdf">https://safesupportivelearning.ed.gov/sites/default/files/EDSCLS_Questionnaires_112_017.pdf</a>
	Adults working at this school treat all students respectfully.		
	People of different cultural backgrounds, races or ethnicities get along well at		
	All students are treated the same, regardless of whether their parents are rich		
	Boys and girls are treated equally well.		
Engagement	How excited are you about going to your classes?	Not at all excited to Extremely excited	The teacher-student relationship scale, school safety scale, sense of belonging scale, learning strategies scale, student engagement scale, and rigorous expectations scale were developed by Panorama Education.  Link to source: <a href="https://www.panoramaed.com/panorama-student-survey">https://www.panoramaed.com/panorama-student-survey</a>
	How often do you get so focused on activities in your classes that you lose track of time?	Almost never to Almost always	
	In your classes, how eager are you to participate?	Not at all eager to Extremely eager	
	When you are not in school, how often do you talk about ideas from your classes?	Almost never to Almost always	
	Overall, how interested are you in your classes?	Not at all interested to Extremely interested	
Learning Strategies	When you get stuck while learning something new, how likely are you to try a different strategy?	Not at all likely to Extremely likely	
	How confident are you that you can choose an effective strategy to get your schoolwork done well?	Not at all confident to Extremely confident	

	Before you start on a challenging project, how often do you think about the best way to approach the project?	Almost never to Almost always	
	Overall, how well do your learning strategies help you learn more effectively?	Not well at all to Extremely well	
	How often do you use strategies to learn more effectively?	Almost never to Almost always	
Rigorous Expectations	How often do your teachers make you explain your answers?	Almost never to Almost always	
	When you feel like giving up on a difficult task, how likely is it that your teachers will make you keep trying?	Not at all likely to Extremely likely	
	How much do your teachers encourage you to do your best?	Do not encourage me at all to Encourage me a tremendous amount	
	How often do your teachers take time to make sure you understand the material?	Almost never to Almost always	
	Overall, how high are your teachers' expectations of you?	Not high at all to Extremely high	
School Safety <i>(reverse coded)</i>	How often are people disrespectful to others at your school?	Almost never to Almost always	
	How often do students get into physical fights at your school?	Almost never to Almost always	
	How likely is it that someone from your school will bully you online?	Not at all likely to Extremely likely	
	How often do you worry about violence at your school?	Almost never to Almost always	
	If a student is bullied in school, how difficult is it for him/her to get help from an adult?	Not at all difficult to Extremely difficult	
	At your school, how unfairly do the adults treat the students?	Not at all unfairly to Extremely unfairly	
Sense of Belonging	How well do people at your school understand you as a person?	Do not understand at all to Completely Understand	

	How connected do you feel to the adults at your school?	Not at all connected to Extremely connected	
	How much respect do students in your school show you?	No respect at all to A tremendous amount of respect	
	How much do you matter to others at this school?	Do not matter at all to Matter a tremendous amount	
	Overall, how much do you feel like you belong at your school?	Do not belong at all to Completely belong	
Teacher-Student Relationship	How many of your teachers are respectful towards you?	None of my teachers to All of my teachers	
	If you walked into class upset, how many of your teachers would be concerned?		
	If you came back to visit class three years from now, how many of your teachers		
	When your teachers ask how you are doing, how many of them are really interested in your answer?		
	How many of your teachers would you be excited to have again in the future?		

## Appendix B

**Table B1. Spring 2017 Student SEC Summary Descriptive Statistics by Subgroup**

	%	Growth Curiosity	Growth Mindset	Perseverance	Self- Awareness	Self- Efficacy	Self- Regulation	Social Awareness
Overall	n=3,028	3.21 (n=3,015)	3.79 (n=3,012)	3.58 (n=3,011)	3.65 (n=3,012)	3.45 (n=3,018)	3.90 (n=3,027)	3.77 (n=3,016)
Female	49%	3.15 (n=1,483)	3.84 (n=1,482)	3.57 (n=1,482)	3.59 (n=1,482)	3.35 (n=1,486)	3.98 (n=1,491)	3.84 (n=1,486)
Male	51%	3.27 (n=1,531)	3.74 (n=1,529)	3.59 (n=1,528)	3.71 (n=1,529)	3.53 (n=1,531)	3.83 (n=1,535)	3.71 (n=1,529)
Latino	53%	3.14 (n=1,596)	3.72 (n=1,596)	3.55 (n=1,596)	3.57 (n=1,595)	3.28 (n=1,598)	3.88 (n=1,599)	3.75 (n=1,598)
Black	18%	3.32 (n=541)	3.68 (n=541)	3.60 (n=540)	3.73 (n=541)	3.55 (n=542)	3.77 (n=547)	3.66 (n=541)
Asian	3%	3.46 (n=82)	4.13 (n=82)	3.75 (n=82)	3.83 (n=82)	3.82 (n=82)	4.11 (n=82)	4.05 (n=82)
Two or More Races	1%	3.14 (n=37)	3.81 (n=37)	3.59 (n=37)	3.79 (n=37)	3.63 (n=38)	4.00 (n=38)	3.81 (n=37)
White	21%	3.30 (n=633)	4.03 (n=631)	3.65 (n=631)	3.80 (n=631)	3.77 (n=633)	4.07 (n=634)	3.93 (n=632)
Other	0.53%	3.21 (n=16)	3.80 (n=16)	3.53 (n=16)	3.91 (n=16)	3.31 (n=16)	4.00 (n=16)	3.86 (n=16)
MENA	3%	3.02 (n=87)	3.73 (n=86)	3.37 (n=86)	3.32 (n=87)	3.16 (n=86)	3.55 (n=88)	3.63 (n=87)
ELL	18%	3.16 (n=520)	3.46 (n=520)	3.46 (n=520)	3.53 (n=520)	3.21 (n=521)	3.76 (n=522)	3.64 (n=521)
Non-ELL	82%	3.22 (n=2,392)	3.86 (n=2,389)	3.60 (n=2,389)	3.67 (n=2,389)	3.48 (n=2,394)	3.94 (n=2,401)	3.80 (n=2,394)
FRPL	64%	3.19 (n=1,780)	3.72 (n=1,777)	3.56 (n=1,776)	3.61 (n=1,778)	3.35 (n=1,782)	3.88 (n=1,788)	3.75 (n=1,780)
Non-FRPL	36%	3.30 (n=1,004)	3.97 (n=1,003)	3.64 (n=1,003)	3.76 (n=1,003)	3.70 (n=1,004)	4.01 (1,007)	3.88 (1,005)
SPED	8%	3.23 (n=237)	3.40 (n=237)	3.45 (n=237)	3.53 (n=236)	3.30 (n=239)	3.66 (n=239)	3.53 (n=238)
Non-SPED	92%	3.21 (n=2,775)	3.82 (n=2,772)	3.59 (n=2,771)	3.66 (n=2,773)	3.46 (2,776)	3.92 (n=2,785)	3.79 (n=2,775)

**Table B2. Spring 2017 Student CC Summary Descriptive Statistics by Subgroup**

	%	Cultural & Linguistic Competence	Engagement	Learning Strategies	Rigorous Expectations	School Safety	Sense of Belonging	Teacher-Student Relationship
Overall	n=3,092	3.67 (n=3,086)	2.90 (n=3,074)	3.51 (n=3,078)	3.89 (n=3,080)	3.80 (n=3,083)	3.26 (n=3,083)	3.52 (n=3,085)
Female	50%	3.69 (n=1,530)	2.89 (n=1,525)	3.53 (n=1,528)	3.91 (n=1,529)	3.82 (n=1,529)	3.24 (n=1,529)	3.51 (n=1,528)
Male	50%	3.66 (n=1,555)	2.91 (n=1,548)	3.50 (n=1,549)	3.87 (n=1,550)	3.80 (n=1,553)	3.29 (n=1,553)	3.54 (n=1,556)
Latino	52%	3.70 (n=1,604)	2.83 (n=1,600)	3.46 (n=1,603)	3.85 (n=1,605)	3.85 (n=1,605)	3.21 (n=1,605)	3.43 (n=1,604)
Black	18%	3.49 (n=555)	3.00 (n=552)	3.51 (n=552)	3.87 (n=553)	3.60 (n=554)	3.24 (n=554)	3.53 (n=555)
Asian	3%	3.88 (n=88)	3.18 (n=87)	3.77 (n=87)	4.15 (n=87)	3.91 (n=87)	3.54 (n=87)	4.02 (n=87)
Two or More Races	1%	3.51 (n=39)	3.03 (n=38)	3.70 (n=38)	4.06 (n=38)	3.81 (n=38)	3.48 (n=38)	3.74 (n=39)
White	22%	3.79 (n=672)	2.98 (n=669)	3.62 (n=670)	3.99 (n=669)	3.89 (n=671)	3.39 (n=671)	3.75 (n=672)
Other	0.49%	3.58 (n=15)	2.64 (n=15)	3.43 (n=15)	3.64 (n=15)	3.94 (n=15)	2.66 (n=15)	3.21 (n=15)
MENA	3%	3.25 (n=91)	2.53 (n=91)	3.42 (n=91)	3.44 (n=91)	3.56 (n=91)	3.11 (n=91)	3.03 (n=91)
ELL	18%	3.66 (n=539)	2.97 (n=537)	3.44 (n=539)	3.80 (n=539)	3.69 (n=539)	3.32 (n=539)	3.56 (n=539)
Non-ELL	82%	3.67 (n=2,334)	2.87 (n=2,330)	3.52 (n=2,329)	3.90 (n=2,330)	3.83 (n=2,332)	3.24 (n=2,332)	3.49 (n=2,334)
FRPL	64%	3.65 (n=1,814)	2.91 (n=1,812)	3.51 (n=1,814)	3.89 (n=1,814)	3.81 (n=1,815)	3.26 (n=1,815)	3.49 (n=1,815)
Non-FRPL	36%	3.72 (n=1,035)	2.95 (n=1,024)	3.60 (n=1,026)	3.98 (n=1,028)	3.80 (n=1,030)	3.32 (n=1,030)	3.65 (n=1,032)
SPED	9%	3.53 (n=264)	2.95 (n=262)	3.37 (n=262)	3.70 (n=261)	3.61 (n=263)	3.23 (n=263)	3.62 (n=263)
Non-SPED	91%	3.69 (n=2,819)	2.89 (n=2,810)	3.53 (n=2,814)	3.90 (n=2,817)	3.82 (n=2,818)	3.26 (n=2,818)	3.51 (n=2,820)

### Appendix C

**Table C1. Cronbach's Alpha by Grade: Student-Reported SEC Constructs (Spring 2017)**

Grade	Constructs						
	Curiosity	Growth Mindset	Perseverance	Self-Awareness	Self-Efficacy	Self-Regulation	Social Awareness
Grade 4	0.84	0.70	0.74	0.73	0.85	0.84	0.85
Grade 5	0.86	0.78	0.79	0.71	0.87	0.87	0.87
Grade 6	0.89	0.80	0.84	0.80	0.92	0.92	0.91
Grade 7	0.89	0.79	0.83	0.76	0.90	0.89	0.87
Grade 8	0.89	0.79	0.84	0.76	0.89	0.89	0.88
Grade 9	0.89	0.79	0.84	0.77	0.91	0.88	0.89
Grade 10	0.87	0.82	0.83	0.72	0.90	0.89	0.88
Grade 11	0.87	0.82	0.79	0.71	0.88	0.81	0.88
Grade 12	0.89	0.90	0.87	0.83	0.91	0.90	0.92

**Table C2. Cronbach's Alpha by Grade: Student-Reported CC Constructs (Spring 2017)**

Grade	Constructs						
	Cultural & Linguistic Competence	Engagement	Learning Strategies	Rigorous Expectations	School Safety	Sense of Belonging	Teacher-Student Relationship
Grade 4	0.66	0.77	0.73	0.74	0.71	0.81	0.83
Grade 5	0.77	0.84	0.79	0.78	0.71	0.79	0.84
Grade 6	0.80	0.83	0.86	0.85	0.76	0.84	0.89
Grade 7	0.81	0.84	0.83	0.84	0.74	0.84	0.88
Grade 8	0.79	0.85	0.86	0.83	0.68	0.83	0.86
Grade 9	0.82	0.85	0.87	0.84	0.72	0.81	0.88
Grade 10	0.84	0.86	0.88	0.83	0.72	0.81	0.85
Grade 11	0.80	0.81	0.84	0.84	0.70	0.76	0.84
Grade 12	0.82	0.89	0.90	0.90	0.79	0.86	0.91