Mosaic™ by ACT® Social Emotional Learning Assessment
Technical Manual: Elementary School Assessment

February 2022
Version 2022.1

Mosaic™ by ACT®
Social Emotional Learning

Commitment to Fair Testing

ACT endorses and is committed to complying with the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). ACT also endorses the Code of Fair Testing Practices in Education (Joint Committee on Testing Practices, 2004), which is a statement of the obligations to test takers of those who develop, administer, or use educational tests and test data in the following four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. ACT endorses and is committed to complying with the Code of Professional Responsibilities in Educational Measurement (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), which is a statement of professional responsibilities for those involved with various aspects of assessments, including development, marketing, interpretation, and use.

We encourage individuals who want more detailed information on a topic discussed in this manual, or on a related topic, to contact ACT.

Table of Contents

List of Tables

List of Figures

Overview of the Technical Manual

The purpose of this technical manual is to provide an overview of the Mosaic™ by ACT® Social Emotional Learning assessment for elementary school students. This technical manual contains an overview of the skills assessed by the assessment, the various item types used, and a description of the procedure used to incorporate the item types into a unified score. In addition, a brief description of the pilot study is provided, along with evidence of reliability and validity from the study and a description of the process used to create the final operational test form. Below is an executive summary of main points and findings, which are followed by more detail in subsequent sections.

- The Mosaic™ by ACT® Social Emotional Learning assessment for elementary school students assesses three social and emotional skills in third through fifth-grade students: Sustaining Effort, Getting Along with Others, and Maintaining Composure. Two dimensions of School Climate are also assessed: Relationships with Teachers and School Safety.

- Three social and emotional skills are assessed with two item types: image-based Likert items and situational judgment tests (SJT), which are combined to yield one unified score per social and emotional skill. School Climate items are assessed using self-report Likert items.

- Users of the assessment are provided with a roster report containing individual student data, an aggregate school-level report, and individual reports for parents/guardians based on each student's results.

- The final assessment form was developed based on the results of three separate pilot studies conducted between 2018 and 2020, each of which tested different combinations and formats of item types. The final assessment contains 18 image-based Likert items, 6 situational judgment test items, 12 self-report Likert items, and 8 additional demographic items. It can be completed by students online in 15–20 minutes.

- Operational assessment items demonstrated evidence of reliability and validity. Average Cronbach's alpha scores are α = .78 and α = .73 for social and emotional and school climate scales, respectively. Evidence was also found for convergent, discriminant, and criterion validity.

Mosaic by ACT Social Emotional Learning Assessment: Elementary School Overview

The Mosaic by ACT Social Emotional Learning assessments are designed to assess the social and emotional skills of students in elementary (grades 3–5), middle (grades 6–8), and high school (grades 9–12). Each assessment also measures students' perceptions of school climate. The elementary assessment is approximately 15–20 minutes in length and can be taken by students online. After completing the assessment, students receive scores on each of the three social and emotional skills and two school climate dimensions assessed. After students complete the assessment, reports are available that capture individual-level student data (roster report), aggregated data across all students within a school who took the assessment (school aggregate report), and individualized student results, with reports designed for families (individual student report). All reports include social and emotional skill scores, and school-level reports (roster and school aggregate) contain information on students' perceptions of school climate.

Assessment Framework

Social and emotional skills can be defined as interpersonal, self-regulatory, and task-related behaviors that are important for adaptation to and successful performance in educational and workplace settings (Casillas, Way, & Burrus, 2015). The skills are distinct from cognitive factors, or intelligence, and are commonly referred to as dispositions, psychosocial skills, life skills, personality, personal skills, and character traits, in addition to the "social and emotional skills" label (Kyllonen, Lipnevich, Burrus, & Roberts, 2014). Social and emotional skills have been shown to be predictive of success in school (e.g., Poropat, 2009), success at work (e.g., Hough & Oswald, 2008), well-being (e.g., Steel, Schmidt, & Shultz, 2008), and several other important life outcomes (e.g., Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Furthermore, a growing literature suggests social and emotional learning (SEL) interventions can effectively enhance students' social and emotional skills, promote positive behavior, and improve academic performance (Corcoran, Cheung, Kim, & Xie, 2018; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Taylor, Oberle, Durlak, & Weissberg, 2017).

Nearly all social and emotional skills can be crosswalked to the Behavioral Skills portion of the ACT® Holistic Framework® (ACT, 2015). The organization of these skills is based on the structure of the Big Five framework (Roberts, Martin, & Olaru, 2015). The Big Five framework, stemming from the field of personality psychology, has been increasingly recognized as a universal framework through which to organize social and emotional skills (e.g., Abrahams et al., 2019; Chernyshenko, Kankaraš, & Drasgow, 2018; Kyllonen et al., 2014; Primi, Santos, John, & De Fruyt, 2016; Soto, Napolitano, &

Roberts, 2021; Walton, Murano, Burrus, & Casillas, 2021). This is due to over 50 years of empirical support documenting relationships between the five factors and critical education and life outcomes (e.g., Barrick, Mount, & Judge, 2001; McCrae & Costa, 1996; Poropat, 2009), cross-cultural replicability (McCrae & Terracciano, 2005; Schmitt, Allik, McCrae, & Benet-Martínez, 2007), and well-documented evidence that skills are amenable to change throughout the lifespan (Hudson, Briley, Chopik, & Derringer, 2019; Roberts, Walton, & Viechtbauer, 2006; Roberts, Luo, Briley, Chow, Su, & Hill, 2017). Furthermore, the Big Five structure replicates with children as young as three years old (Halverson et al., 2003; Mervielde & DeFruyt, 1999; Tackett et al., 2012).

While the Mosaic SEL middle and high school assessments measure five social and emotional skills, each directly aligned to a Big Five factor, the elementary assessment measures three key social and emotional skills. This design decision stemmed from a convergence of theoretical support, empirical support from pilot study data (see p. 16), and customer feedback (see p. 15–16).

In childhood, social and emotional skills develop in tandem with other social and cognitive capabilities. Throughout childhood, there are developmental milestones and shifting social contexts that dictate which skills are most developmentally relevant for student success at each stage (Brackett, Elbertson, & Rivers, 2015). During the elementary years, children grow social networks and have more complex social interactions with peers compared to early childhood years. Focal developmental milestones include initiating and maintaining friendships with peers, becoming aware of context-dependent scenarios in which they should either express or manage their emotions, and making responsible decisions regarding school behavior and interactions with peers and adults (Denham, 2015). A review of SEL programming for elementary-aged students showed programs focused on interpersonal skills were the most effective in improving students' social and emotional skills (Rimm-Kauffman & Hulleman, 2015). Meta-analytic evidence, which combines large samples of primary studies, also shows that conscientiousness ($r =$ .28), agreeableness ($r = $.30), and emotional stability ($r = $.20) all relate to academic performance in primary school years (Poropat, 2009). Emotional stability, agreeableness, and conscientiousness have also been associated with positive attitudes toward school (Heaven, Mak, Barry, & Ciarrochi, 2002), overall well-being, life satisfaction, and mental health (Chernyshenko et al., 2018). Given the developmental relevance of these three skills supported by previous literature, as well as convergence with empirical evidence and customer feedback from the pilot studies detailed in later sections, Sustaining Effort, Getting Along with Others, and Maintaining Composure are the three social and emotional skills measured in the elementary school assessment. Definitions for each social and emotional skill are provided below in Table 1.

Alignment to CASEL's SEL Competencies

Each social and emotional skill as defined by the ACT Holistic Framework also aligns to the Collaborative for Academic, Social, and Emotional Learning (CASEL) SEL competencies. Alignment to each CASEL competency was agreed upon by five subject matter experts (see Walton, Burrus, Anguiano-Carrasco, Way, & Murano, 2019, for full methodology). Table 1 provides descriptions of each Mosaic by ACT skill measured in the elementary school assessment and their alignment to corresponding Big Five factor and CASEL competency.

Table 1. Mosaic Social Emotional Learning Elementary School Assessment Skills, Definitions, and Alignment to CASEL Competencies

Mosaic Social Emotional Skill	Big Five Alignment	CASEL Alignment	Skill Definition
Sustaining Effort	Conscientiousness	Self-Management Responsible Decision-Making	How a student's actions demonstrate diligence, effort, organization, self-control, and compliance with the rules.
Getting Along with Others	Agreeableness	Social Awareness Relationship Skills	How a student's actions demonstrate positive interactions and cooperation with others, kindness, friendliness, and tactfulness.
Maintaining Composure	Emotional Stability	Self-Awareness Self-Management	How a student's actions demonstrate relative calmness, serenity, and the ability to manage emotions effectively.

School Climate

In addition to social and emotional skills, the assessment also measures two school climate dimensions: Relationships with Teachers and School Safety. The relationships dimension focuses mainly on student-teacher relationships. These relationships tend to be related to frequency of behavioral problems (Gregory & Cornell, 2009) and engagement in the classroom (Skinner & Belmont, 1993), with more positive climate being associated with fewer behavioral problems and more classroom engagement. Feelings of safety at school have been shown to promote learning (Devine & Cohen, 2007), whereas feeling unsafe at school is related to higher levels of absenteeism and lower levels of academic achievement (ACT, 2016). Importantly, these dimensions are also related to social and emotional skills, with a recent study finding that relationships and safety relate to student motivation, self-regulation, and social engagement (Allen, Way, & Casillas, 2019). Positive school climate and positive social and emotional skills have a bidirectional relationship, with one supporting the other (Osher & Berg, 2017).

Assessment Overview

The elementary school assessment measures three social and emotional skills (Sustaining Effort, Getting Along with Others, and Maintaining Composure) and two dimensions of school climate (Relationships with Teachers and School Safety Climate) using two item types: Likert items and situational judgment test items. The purpose of using multiple item types is to reduce the impact of biases or other shortcomings associated with the use of any single item type. For example, while Likert items are efficient to administer, they are also susceptible to various response biases such as reference bias (see Ziegler, MacCann, & Roberts, 2012). Alternatively, while situational judgment test items eliminate the risk of reference bias and are difficult to fake, they increase test time and reading load (e.g., Hooper, Cullen, & Sackett, 2006; Lipnevich et al., 2013). For more detail on advantages and disadvantages for each item type and evidence supporting the validity of a unified scoring approach, see ACT (2021). All items are administered in an online assessment and are student self-report.

Likert Items

Likert items measuring social and emotional skills are image-based. In these items, students are presented with an image and accompanying text that describe behaviors related to each social and emotional skill. Students use a four-point scale to rate how well each image describes them and how they typically act. Likert items measuring school climate are text-based, and students rate their level of agreement with each statement on a four-point scale.

Figure 1. Sample Image-Based Likert Item

I can get along with other kids.

○ Not like me at all

○ Kind of like me

○ Mostly like me

○ A lot like me

Situational Judgment Test Items

Situational judgment test items are text-based. In these items, students are presented with scenarios and three possible behavioral responses to those scenarios. Students use a four-point scale to rate how likely they are to engage in each behavioral response.

Figure 2. Sample Situational Judgment Test Item

A new student just moved to your town and joins your class in the middle of the year. What would you do?

Ask if he wants to play with you and your friends at recess.	○ I would never do that	○ I probably wouldn't do that	○ I might do that	○ I would definitely do that
Don't talk to him because you have a lot of friends in your class already.	○ I would never do that	○ I probably wouldn't do that	○ I might do that	○ I would definitely do that
Be friendly to him to make him feel welcome.	○ I would never do that	○ I probably wouldn't do that	○ I might do that	○ I would definitely do that

Assessment Reports

Users receive three different reports: a school aggregate report, a roster report, and an individual report for each student designed for families.

The school aggregate report contains aggregated data on all students within the school who took the assessment. Scores are reported for each social and emotional skill and for each school climate scale and item. This report is useful in allowing schools to compare their results with those of other schools, in comparing subgroup results across grades, and in tracking a school's overall assessment performance over time. It also includes lesson content designed to help students build their social and emotional skills.

The roster report contains individual-level data on each student within the school who took the assessment. Social and emotional skills are reported both as scores on a 1 (Developing) to 4 (Mastering) point scale and percentile scores, and school climate scores are reported as scale scores on a 1 (minimum) to 4 (maximum) point scale. The roster report also includes student identifying and demographic information, student self-reported academic performance in math and English Language Arts (ELA), and students' perceptions of their self-efficacy in math and ELA.

Individual reports are also provided for each student who completed the assessment. These reports are designed for family members and contain information and students' level scores on each social and emotional skill. They also include lesson content designed to help students build their social and emotional skills.

Unified Score

For each social and emotional skill, a unified score is computed using the Likert and situational judgment test responses per skill. Mean scores for each item in each method are computed, z-standardized, and averaged to create a unified score for each skill.

Item Development

Three separate pilot studies were conducted in the development and validation of the elementary assessment items. This section describes the initial development process and briefly summarizes high-level findings from each. For more detail on item content and results from pilot studies 1 and 2, see Murano et al. (2020).

Initial Item Development

To generate the initial pool of items for elementary school students, subject matter experts drafted adjectives aligned to definitions of five social and emotional skills (Sustaining Effort, Getting Along with Others, Maintaining Composure, Keeping an Open Mind, and Social Connection). Eight adjectives were written per skill for a total of 40 image-based Likert items. A professional artist also partnered to develop images to accompany each adjective. Images were created based on the item text and were intentionally designed to be gender and race/ethnicity neutral (see page 11 for an example). Images were reviewed by subject matter experts and revised when necessary. Once Likert items were complete and paired with revised images, 30 of the 40 were arranged into forced choice triads with three response options in each triad, one of which was negatively keyed (see Murano et al., 2020, for more detail on forced choice items). Item writers also developed 20 situational judgment test items, each containing a scenario and five behavioral response options. Cognitive labs were conducted with nine elementary-aged students in Iowa City to test for understanding and clarity of item content and instructions prior to the first pilot study.

Item Selection and Revisions

The initial item pool was administered to 1,364 students grades 3–5 from 12 elementary schools across the United States. Prior to analyses, instances of low-quality responses were removed. Cases were excluded if they demonstrated any of the following response patterns associated with inattentive responding: excessive missing data (> 20%), response time shorter than half the median testing time of the student's grade, variance < 0.1 on Likert or SJT items, or identical FC response patterns for all FC items. Of 1,364 original cases in the data set, 1,047 were used for analyses (n = 342 third grade, n = 411 fourth grade, and n = 293 fifth-grade students).

For the image-based Likert items, alphas for 8-item scales ranged from α = .47–.78 (average α = .63). A confirmatory five-factor model was fit to the Likert data using weighted least squares estimation, with fit statistics as follows: *RMSEA*=0.085 (CI: 0.082–0.087), *CFI*=0.883, and *TLI*=0.871. The same analytic procedure was used for the situational judgment test items. Alphas ranged from α = .17–.80 (average α = .59), and a confirmatory factor model showed fairly poor fit: *RMSEA*=0.083 (CI: 0.081–0.084), *CFI*=0.718, and *TLI*=0.689. For the forced choice items, a Thurstonian IRT model was fit in an attempt to generate normative scores (see Brown & Maydeu-Olivares, 2013, for more detail). However, the model failed to reach convergence. Ipsative scores were generated instead with alphas ranging from α = .24–.59

(average α = .43). Convergent, discriminant, test-criterion, and incremental validity were also evaluated (see Murano et al., 2020, for more detail).

Based on reliability and validity evidence from the first study, item revisions were made prior to the second pilot study. Most notably, Likert items were re-written as statements instead of adjectives to create more concrete, less abstract items for younger students. Content revisions were made to items with poor loadings that demonstrated low internal consistency or contained double negatives in the response set (e.g., needing to rank "not shy" on a scale from *not like me at all* to *a lot like me*). Situational judgment tests were revised to include only three possible behavioral responses instead of five. This stemmed from feedback from pilot participants requesting a shorter test time. The revised item pool was piloted with 925 students grades 3–5 from seven elementary schools in the Midwest. The same data quality screening procedures were used as in Pilot 1, resulting in 826 complete cases (*n* = 208 third grade, *n* = 323 fourth grade, and *n* = 295 fifth grade students).

All analytic procedures were identical to those conducted for Pilot 1. For the image-based Likert items written as sentences, alphas for 8-item scales ranged from α = .72–.80 (average α = .75). A confirmatory five-factor model was fit to the Likert data using weighted least squares estimation, with improved fit statistics as follows: *RMSEA*=0.078 (CI: 0.075–0.081), *CFI*=0.913, and *TLI*=0.904. For situational judgment tests, alphas ranged from α = .27–.62 (average α = .51), and a confirmatory factor model again showed improved fit: *RMSEA*=0.080 (CI: 0.077–0.083), *CFI*=0.809, and *TLI*=0.801. For the forced choice items, the Thurstonian IRT model again failed to reach convergence. Ipsative scores were generated instead with alphas ranging from α = .39–.58 (average α = .51). Convergent, discriminant, test-criterion, and incremental validity were also evaluated (see Murano et al., 2020, for more detail).

Final Item Set Development

Based on the results from Pilot 2, further revisions were made to the item pool. The sentence format was retained, given stronger reliability and validity evidence from Pilot 1 to Pilot 2. Most notably, forced choice triads were rearranged to only contain positively keyed items and matched on social desirability ratings, rather than each triad containing one negatively keyed statement. Likert and situational judgment test items with poor loadings were revised based on content. A modified version of the Relationships (12 items) and Safety (11 items) school climate scales from ACT® Engage® were also included (for information on the development of those scales, see ACT, 2016). The full item pool was piloted with 1,439 participants in 23 schools across the United States.

In addition to piloting the item pool, a survey was also completed by 33 school staff members who administered the assessment to students. Participants were asked to rate their students' testing experience, any difficulties, and engagement with

assessment items. While the majority of participants responded favorably regarding students' willingness to take the assessment (86.21% agreement), engagement in the assessment (67.86% agreement), the assessment's age and context-relevance (92.85% agreement), and appropriate reading level (68.96%), users found the assessment to be too long for students (62.07% agreement), with multiple users reporting this concern in open-ended responses. We also asked users to rate students' understanding and engagement of each item type. Figure 1 displays average participant responses for student engagement with each item type. Items were on a 1 (strongly disagree) to 4 (strongly agree) scale. Overall, students had the least positive experience answering force choice items.

Figure 3. Student Experience with Assessment Items

- My students understood how to answer these items
- My student found these items engaging
- My student had trouble answering these items
- I would like to see these items included in future SEL assessments
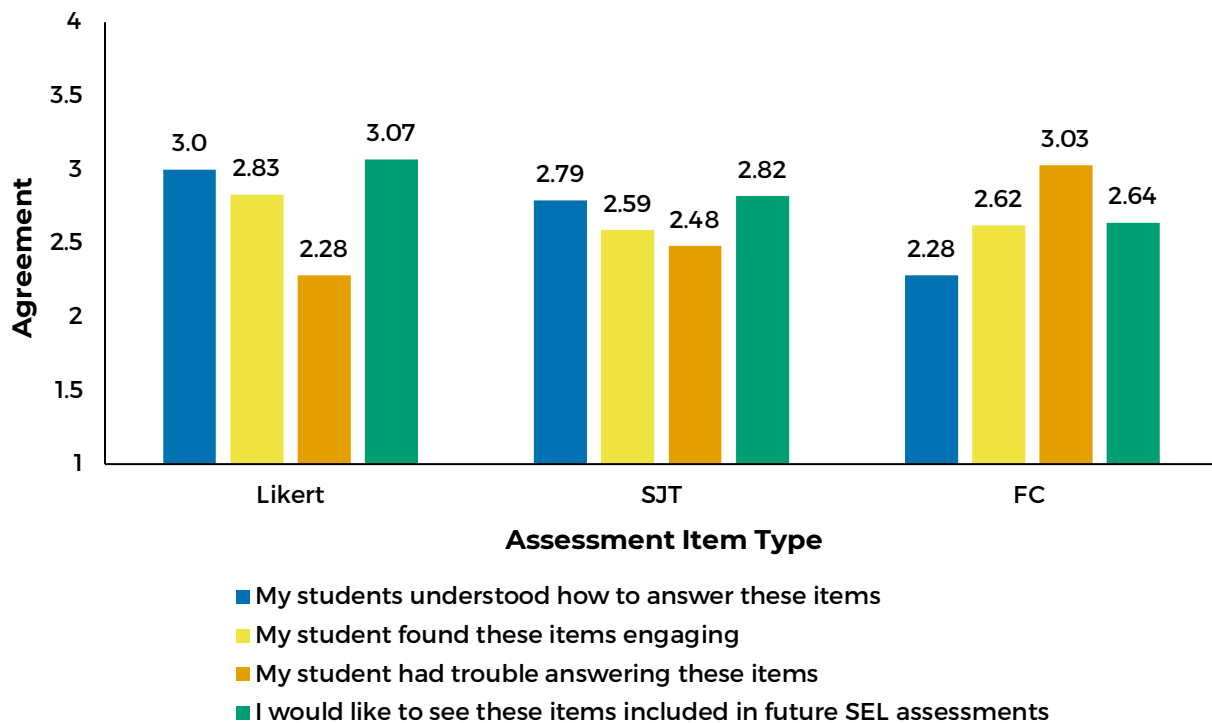
Note. Reponses are on a 1 (strongly disagree) to 4 (strongly agree) scale with higher responses indicating higher agreement.

Users were also asked to rank the importance of each of the five social and emotional skills. Getting Along with Others was the skill most frequently ranked as being the most important by users (39.39%), followed by Sustaining Effort (27.27%) and Maintaining Composure (21.21%). Figure 2 shows the counts of users who rated each skill in each position (1 on the x-axis is most important, 5 is least important).

Figure 4. Counts of Skill Importance Ratings
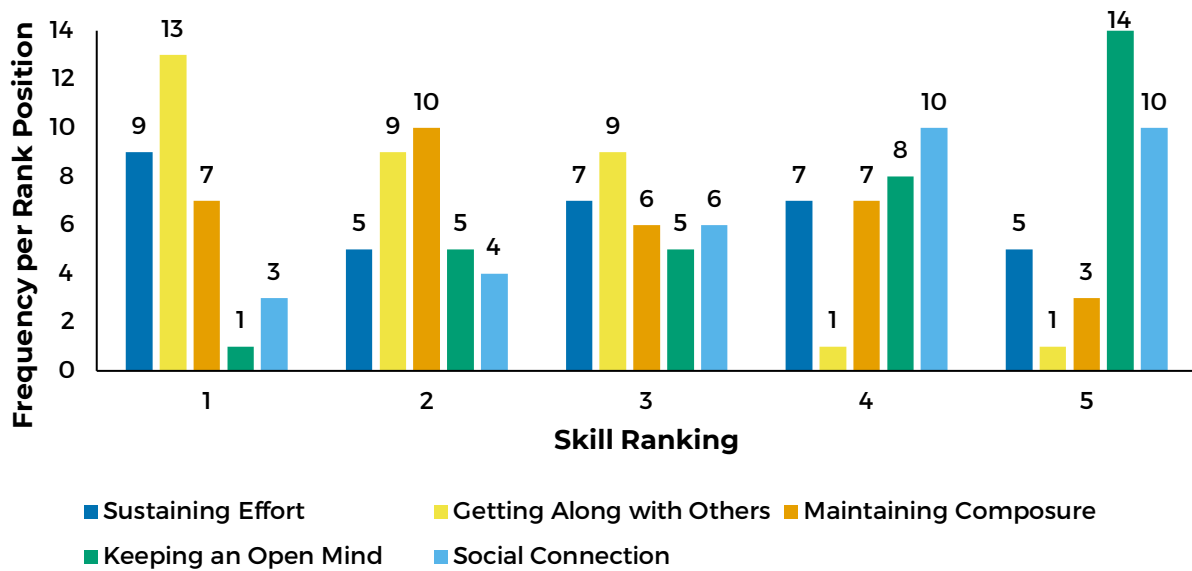
Note. On the x-axis, 1 corresponds to the most important skill, and 5 corresponds to the least important skill. Counts of respondents who ranked each skill in each respective category are reported.

Based on user feedback, we examined the pilot data to determine if the assessment could be shortened to meet user needs. Several design decisions were made based on both data sets. First, the forced choice with all positively keyed items still failed to converge when fit to the Thurstonian IRT model. Based on non-convergence issues and user feedback regarding students' difficulties, we eliminated forced choice items from the final test form. Second, we considered reducing the number of skills measured. When reduced to measuring three social and emotional skills compared to five, CFA model fits improved for both Likert and situational judgment test items (see Table 2). Theoretical support, empirical data from the pilot study, and user feedback therefore supported the decision to measure only three skills on the elementary version of the assessment: Sustaining Effort, Getting Along with Others, and Maintaining Composure.

Table 2. Confirmatory Factor Analysis Fit Statistics for Five-Factor and Three-Factor Models

Model	RMSEA (90% C.I.)	CFI	TLI
Likert (5 factor)	.055 (.052–.057)	.900	.890
Likert (3 factor)	.056 (.055–.060)	.913	.899
SJT (5 factor)	.087 (.084–.090)	.728	.701
SJT (3 factor)	.071 (066–.075)	.776	.743

The final test form contains 18 image-based Likert items (6 per social and emotional skill), 6 situational judgment tests (2 per social and emotional skill), and 12 self-report Likert items (6 per school climate scale). The Flesch-Kinkaid reading level for the final assessment items is 2.3. The following section contains results demonstrating evidence of reliability and validity for the final operational test form.

Field Study Results

Field Study Participants

The assessment was administered to 1,439 students grades 3–5 from 23 elementary schools across the United States. Prior to analyses, instances of low-quality responses were removed. Cases were excluded if they demonstrated any of the following response patterns: excessive missing data (> 20%), response time shorter than half the median testing time of the student's grade, variance < 0.1 on Likert or SJT items, or identical FC response patterns for all FC items. Of 1,439 original cases in the data set, 1,206 were used for analyses. Demographic information of participants is reported in Table 3.

Table 3. Field Study Participant Demographic Information

Demographic Variable	Number of Students	Percentage of Students
Grade		
3rd grade	368	30.5%
4th grade	425	35.2%
5th grade	413	34.2%
Gender		
Female	613	51.1%
Male	543	45.3%
Other	5	<0.1%
Prefer not to respond	39	3.2%
Race/Ethnicity		
American Indian or Alaska Native	59	4.9%
Asian	23	1.9%
Black or African American	187	15.6%
Hispanic, Latinx, or of Spanish Origin	107	8.9%
Native Hawaiian or Other Pacific Islander	11	<0.1%
White	543	45.0%
Two or more races	78	6.5%
Prefer not to respond	188	15.6%
Total	1206	

Note. Reponses for gender status were missing for 6 students and for race/ethnicity for 10 students.

Field Study Validation Measures

In addition to the social and emotional skills and school climate assessment items, participants also completed the Big Five Questionnaire for Children (BFQ-C; Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003), which is a measure of the Big Five personality factors validated for use with children as young as six years old. Thirty of the items with the highest factor loadings (six per social and emotional skill) were included in the field study to evaluate the convergent and discriminant validity of the Mosaic social and emotional skill scales. Each Mosaic skill was expected to correlate most highly with its respective Big Five factor. Participants also completed a four-item scale measuring their attitudes toward school. This scale was adapted from PISA items that measured students' attitudes toward mathematics by changing language referring to "math" to "school" (OECD, 2012). A sample item is: "*I learn things in school that are important.*" Participants also self-reported their perception of their overall academic performance and grades in their ELA and mathematics classes. Academic performance was expected to correlate most highly with Sustaining Effort (Poropat, 2009).

Field Study Results

Evidence for Reliability

Each image-based Likert scale consists of six items and demonstrates acceptable reliability evidence. Cronbach's alpha values are α = .73 (Sustaining Effort), α = .77 (Getting Along with Others), and α = .73 (Maintaining Composure).

Situational judgment test scales consist of six items per skill (the average of three responses per scenario for two scenarios per skill) and demonstrate acceptable reliability evidence. Cronbach's alpha values are α = .61 (Sustaining Effort), α = .72 (Getting Along with Others), and α = .69 (Maintaining Composure). These estimates are particularly notable considering that average internal consistency ratings for SJT scales average alphas of 0.57 (Campion, Ployhart, & MacKenzie, 2014). That is, SJT scales are expected to have lower internal consistency estimates than Likert items due to their complex, multi-dimensional nature.

To create a unified score for each social and emotional skill, scores generated from Likert and situational judgment test scales were z-standardized and averaged. Unified scores demonstrate acceptable reliability evidence with Cronbach's alpha

values of α = .77 (Sustaining Effort), α = .80 (Getting Along with Others), and α = .79 (Maintaining Composure).

Each school climate scale consists of six items and demonstrates acceptable reliability evidence. Cronbach's alpha values are α = .78 (Relationships with Teachers) and α = .68 (School Safety).

Evidence for Validity

We collected two primary sources of validity evidence in the field study: evidence based on internal structure and evidence based on relations to other variables (see AERA, APA, & NCME, 2014, for a full description of sources of validity evidence). A confirmatory three-factor model was fit to the image-based Likert data to assess structural validity using weighted least squares estimation. Fit statistics are as follows: *RMSEA*=0.056 (CI: 0.052–0.060), *CFI*=0.913, and *TLI*=0.899. For the SJT items, fit statistics for a confirmatory three-factor model are as follows: *RMSEA*=0.071 (CI: 0.066–0.075), *CFI*=0.776, and *TLI*=0.743. For the unified score model, fit statistics for a confirmatory three-factor model are as follows: *RMSEA*=0.076 (CI: 0.074–0.078), *CFI*=0.855, and *TLI*=0.845. For the school climate scales, fit statistics for a confirmatory two-factor model are as follows: *RMSEA*=0.055 (CI: 0.051–0.061), *CFI*=0.897, and *TLI*=0.871.

To assess validity evidence based on relations with other variables, we compared unified scores for each social and emotional skill scores to scores obtained from additional outcome measures collected in the field study. The correlation matrix below presents the inter-scale correlation matrix for all unified social and emotional skill scores, the BFQ-C scales, attitudes toward school scale, and self-reported academic performance overall in school.

Table 4. Inter-Scale Correlation Matrix

Scale		1	2	3	4	5	6	7	8	9
Mosaic SEL Skill										
1	Sustaining Effort									
2	Getting Along with Others	0.61								
3	Maintaining Composure	0.60	0.58							
Mosaic School Climate										
4	Relationships with Teachers	0.50	0.51	0.48						
5	School Safety	0.21	0.18	0.23	0.38					
BFQ-C										
6	Conscientiousness	0.64	0.47	0.54	0.45	0.21				
7	Agreeableness	0.55	0.68	0.54	0.49	0.30	0.63			
8	Emotional Stability	0.29	0.26	0.45	0.19	0.08	0.14	0.19		
School Outcomes										
9	Attitude toward school	0.56	0.53	0.48	0.49	0.25	0.52	0.58	0.29	
10	Overall school performance	0.35	0.30	0.28	0.31	0.14	0.27	0.32	0.11	0.32

Note. $N = 1,206$ unless otherwise indicated. BFQ-C Conscientiousness $N = 1,143$, Agreeableness $N = 1,143$, Emotional Stability $N = 1,128$. School attitude $N = 1,063$. Academic performance $N = 945$.

Results reported in the inter-scale matrix demonstrate evidence of convergent validity as each social and emotional skill scales correlated most highly with their respective Big Five factor, with correlations averaging .59. While the emotional stability BFQ-C scale correlated higher with the Maintaining Composure scale (.45) than Sustaining Effort (.29) or Getting Along with Others (.26), the Maintaining Composure scale showed higher correlations with the Conscientiousness and Agreeableness BFQ-C scales (both .54) than Emotional Stability. This could stem from exclusively negatively keyed items in the BFQ-C measure. Though scales are moderately correlated in general, there is also evidence for discriminant validity with off-trait correlations averaging .44.

Results reported in the inter-scale matrix also demonstrate evidence of criterion-related validity. As expected, all social and emotional skills are positively correlated with school climate perceptions, with Relationships with Teachers more highly correlated with skills (average correlation = .50) than School Safety (average correlation = .21). Furthermore, each social and emotional skill shows positive correlations with a positive attitude toward school. As expected, Sustaining Effort is the social and emotional skill most highly correlated with academic performance.

Additional Associations with Academic Performance

Students were asked to self-report their overall academic performance by answering a single item, "How well do you think you are doing in school overall?", with response options from 1 (Not very well) to 4 (Very well). Additionally, students were asked to self-report the grades they received in ELA and math courses. Given the lack of standardization across elementary schools in assigning grades, students were given four categories with multiple labels and asked to choose which best describe the grade metric they received. Categories included *A/4/90–100%/Exceeds Expectations, B/3/80–90%/Meets Expectations, C/2/70–80%/Partially Meets Expectations, D–F/1/60–70%/Does Not Meet Expectations, I don't know,* **and** *I don't receive grades.* Correlations between each metric of academic performance with social and emotional skills and school climate scales are reported in Table 5. Based on previous meta-analytic findings (Poropat, 2009), Sustaining Effort was expected to have the strongest relationship with academic performance, followed by Getting Along with Others. The positive correlations between GPA and school climate scales also replicate prior research suggesting a positive association between academic performance and school climate (Berkowitz et al., 2017).

Table 5. Correlations Between Social and Emotional Skills and School Climate and Academic Performance

| Assessment Scale	Overall Perceived Performance (*n* = 945)	ELA Grades (*n* = 630)	Math Grades (*n* = 716)
Sustaining Effort	.35*	.20*	.22*
Getting Along with Others	.30*	.12*	.13*
Maintaining Composure	.28*	.06	.17*
Relationships with Teachers	.31*	.04	.15*
School Safety	.14*	.04	.07

Note. *p < .05.

Regression models were also fit with social and emotional skills as predictors for each outcome of interest. The social and emotional skill scales accounted for a statistically significant amount of variance in overall student-reported perceived academic performance: R^2 = .14, $F_{(3, 941)}$ = 49.41, and p < .01. The scales also accounted for a small, but significant amount of variance in student-reported ELA grades: R^2 = .05, $F_{(3, 626)}$ = 10.21, and p < .01; and math grades: R^2 = .05, $F_{(3, 712)}$ = 12.81, and p < .01.

School climate scales were also entered into each regression model as second blocks to determine if school climate accounted for incremental variance in perceived academic performance and grades beyond that accounted for by social

and emotional skills. For overall perceived academic performance, adding school climate resulted in a very small, but significant amount of variance: R^2 = .15, $\triangle R^2$ = .01, $F_{(2, 941)}$ = 7.99, and p < .01. School climate scales did not account for any additional variance in ELA grades: R^2 = .05, $\triangle R^2$ < .01, $F_{(2, 624)}$ = 2.03, and p = .13; or math grades: R^2 = .05, $\triangle R^2$ < .01, $F_{(2, 710)}$ = .42, and p = .66. Overall, social and emotional skills accounted for most of the variance seen in academic performance.

Figure 5 illustrates the relationship between social and emotional skills and student-reported perceived academic performance. Students were identified as scoring in the first (bottom 25%), second (second 25%), third (third 25%), or fourth (top 25%) quartile on each skill. The y-axis represents student overall perceived academic performance, with higher values representing stronger perceived academic performance. Students in the lowest quartile showed a perceived performance difference of about 1 full point lower than students in top quartile.

**Figure 5**. Overall Perceived Academic Performance by Social and Emotional Skill Quartile
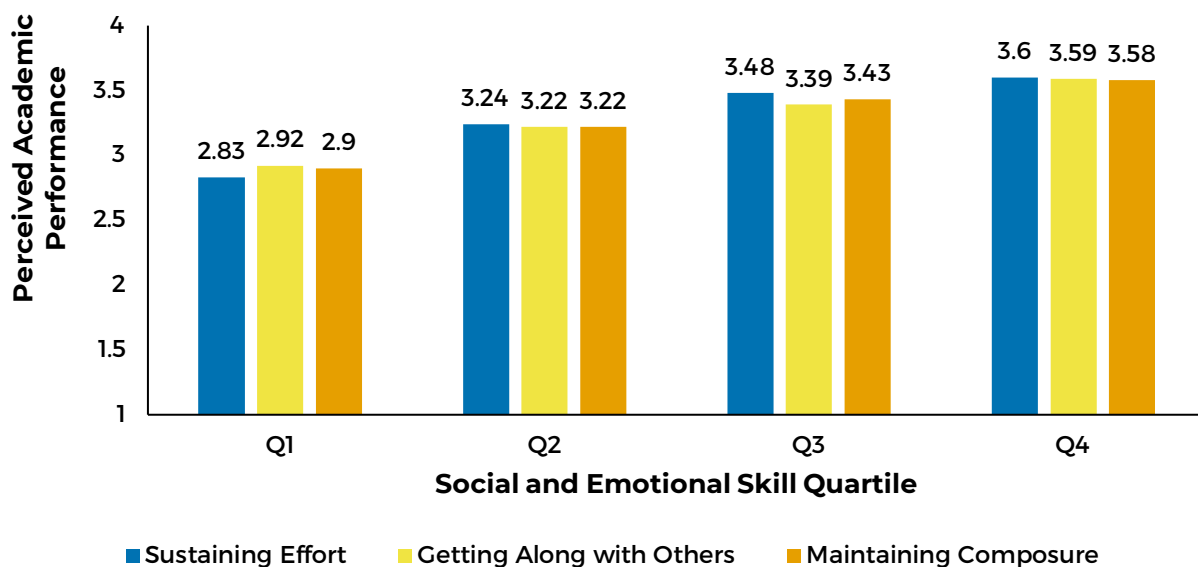
Note. On the y-axis, 1 corresponds to the lowest level of overall perceived academic performance (*not very well*), and 4 corresponds to the highest level of overall perceived academic performance (*very well*). On the x-axis, Q1 refers to the first quartile, Q2 the second quartile, Q3 the third quartile, and Q4 the fourth quartile.

Figure 6 further illustrates the relationship between social and emotional skills and grades in math and ELA. Students responded to a single item for math and a single item for ELA asking them what their grades were in each subject, with the response categories: *A/4/90–100%/Exceeds Expectations, B/3/80–90%/Meets Expectations, C/2/70–80%/Partially Meets Expectations, D–F/1/60–70%/Does Not Meet Expectations, I don't know,* and *I don't receive grades.* **Below, we show the relationship between Sustaining Effort, the skill most strongly related to academic performance, and ELA and math grades. Students are represented by their skill level: Developing (1st–16th percentile), Approaching (17th–50th percentile), Demonstrating (51st–83rd percentile), and Mastering (84th–99th percentile). As shown in the graph, students at the Mastering level are about 50% more likely to receive mostly A grades in their classes compared to students at the Developing level.**

Figure 6. Distributions of Math and ELA Grades by Sustaining Effort Level

Math Grades

ELA Grades
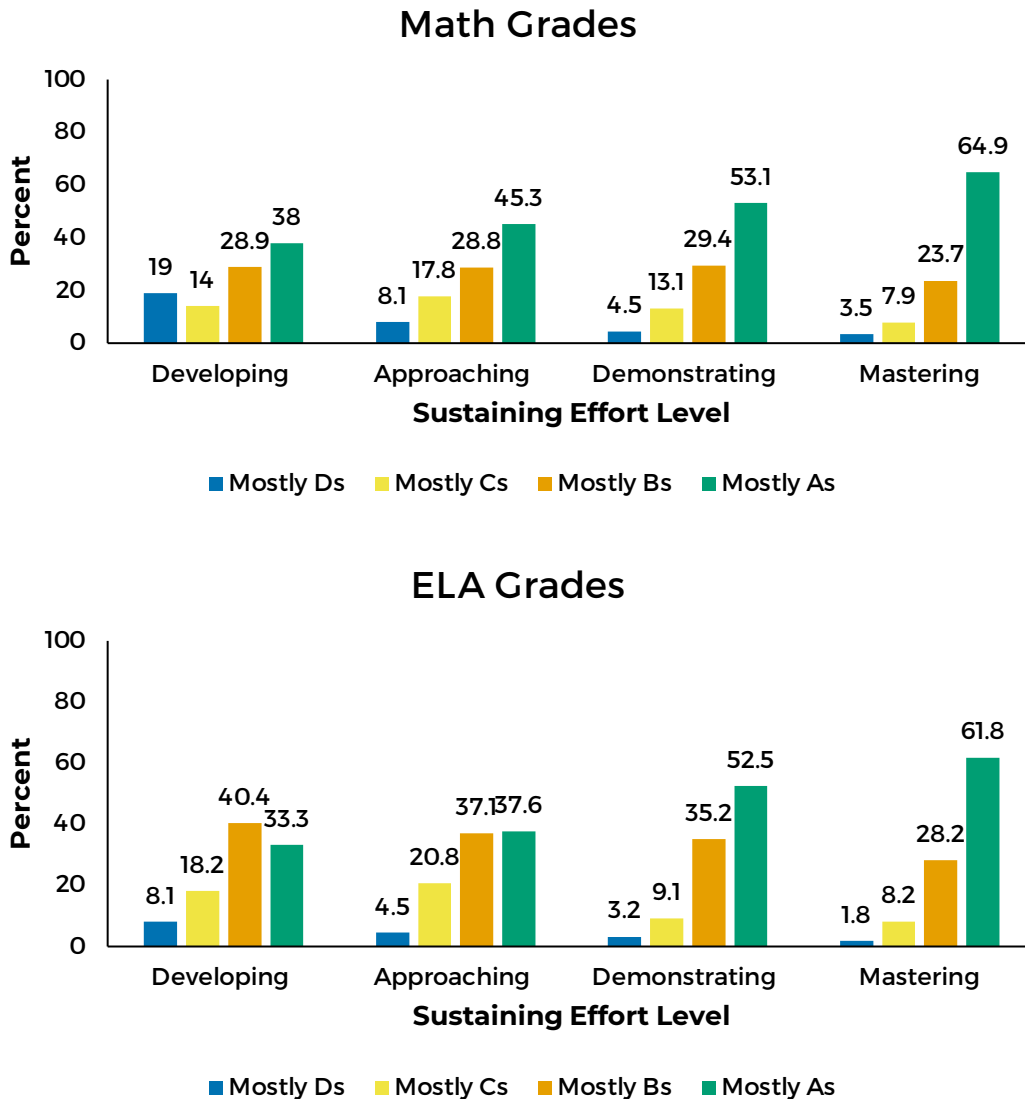
Subgroup Differences

Demographic group differences were evaluated to determine whether there were any significant subgroup differences on the assessment scales. For gender, independent samples t-tests were carried out to compare students self-identified as male vs. female. Students who reported "other" or preferred not to respond were excluded from the analyses given the small sample sizes. Descriptive statistics for each scale by subgroup, as well as t-test results and standardized effects, can be found in Tables 5. For race-ethnicity status, independent samples t-tests were carried out to compare students who identified as White versus students who identified as non-White. The non-White group included any students who identified as a racial or ethnic group other than White or Asian. This included students who identified as American Indian, Black, Hispanic, and multi-racial students. Students who reported "other" or preferred not to respond were excluded from the analyses. Descriptive statistics for each scale by subgroup, as well as t-test results and standardized effects, can be found in Table 6. To evaluate grade differences, one-way ANOVA tests were carried out to compare students across grades on mean scale scores. Alphas are also reported for each separate grade. Table 7 contains scale scores, standard deviations, and alphas by grade.

Gender Differences

Previous literature has shown gender differences emerge between females and males, with females scoring higher on measures of social and emotional skills compared to males from early–middle childhood through early adolescence (e.g., De Bolle, 2015; Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006; Olino, Durbin, Klein, Hayden, & Dyson, 2013). Our findings are largely in line with the body of literature showing subgroup differences favoring females. Table 5 reports all means, standard deviations, t-values, and standardized effect sizes (Cohen's d) for male-female comparisons.

Table 6. Gender Differences on Social and Emotional Skill and School Climate Scales

Assessment Scale	Female[a]		Male[b]		t	d
	M	SD	M	SD		
Sustaining Effort	.11	.82	−0.09	.86	4.00*	.24
Getting Along with Others	.15	.76	−0.13	.88	5.57*	.33
Maintaining Composure	.06	.82	−0.04	.86	1.93	.11
Relationships with Teacher	3.45	.50	3.35	.60	2.93*	.17
School Safety	3.16	.42	3.14	.50	.65	.04

Notes. *p < .05. [a]N = 613. [b]N = 543. Effects are reported in the direction of females (i.e., a positive effect shows that females scored higher than males).

Race/Ethnicity Differences

According to Foldes, Duehr, and Ones's (2008) meta-analytic data, White and Black students generally show negligible to small subgroup differences in social and emotional skills with a few exceptions (exceptions are those with $d > 20$). White students score higher on the emotional stability facet of low anxiety and score higher on global measures of extraversion and the facet of sociability. For the most part, small differences are found between White and Hispanic students also, but Hispanic students score higher on low anxiety (Foldes et al., 2008). In our data, we did see slightly larger effects sizes than those observed in Foldes et al. (2008). However, these effects can still be considered relatively small in magnitude as per Cohen's (1992) seminal guidelines, as well as more recent interpretation guidelines for social and behavioral data (Schäfer & Schwarz, 2019). There were no significant subgroup differences observed in our school climate data. This differed from previous data showing that White students typically report higher climate scores compared to non-White students (Koth et al., 2008; Mitchell et al., 2010).

Table 7. Race/Ethnicity Differences on Social and Emotional Skill and School Climate Scales

Assessment Scale	White[a]		Non-White[b]			
	M	SD	M	SD	t	d
Sustaining Effort	.06	.81	−0.04	.89	1.96*	.13
Getting Along with Others	.10	.78	−0.12	.92	3.99*	.26
Maintaining Composure	.10	.80	−0.09	.88	3.43*	.23
Relationships with Teacher	3.43	.53	3.37	.43	1.80	.12
School Safety	3.18	.43	3.14	.50	1.32	.09

Notes. *$p < .05$. [a]$N = 566$. [b]$N = 442$. Effects are reported in the direction of White students (i.e., a positive effect shows that White students scored higher than non-White students).

Grade-Level Differences

One-way ANOVA models were fit to evaluate the scales at each grade level. For Sustaining Effort, there was a significant difference between groups: $F_{(2,1203)} = 3.12$, $p < .05$. Bonferroni post-hoc analyses showed that fourth-grade students scored significantly higher than fifth-grade students, with no other grade-level comparisons reaching significance. The Maintaining Composure scale also showed differences across grade levels: $F_{(2,1203)} = 4.15$, $p < .05$. Bonferroni post-hoc analyses showed that third-grade students scored significantly higher than fifth-grade students, with no

other grade-level comparisons reaching significance. There were no significant differences between grades for the Getting Along with Others, Relationships with Teachers, and School Safety scales. Table 8 reports means and standard deviations for all scale scores, in addition to scale alphas for the unified scores by grade. Alphas increased with student age, with the exception of the Relationships with Teachers scale, where alpha was lowest for fifth-grade students and highest for those in third.

Table 8. Grade Differences on Social and Emotional Skill and School Climate Scales

	3rd grade[a]			4th grade[b]			5th grade[c]		
	M	SD	α	M	SD	α	M	SD	α
Sustaining Effort	−.01	.89	.74	.08	.81	.75	−.07	.86	.80
Getting Along with Others	.02	.87	.79	.03	.82	.79	−.04	.86	.83
Maintaining Composure	.09	.83	.76	−.01	.87	.80	−.08	.83	.79
Relationships with Teachers	3.40	.56	.78	3.41	.56	.77	3.36	.55	.76
School Safety	3.14	.48	.68	3.15	.45	.65	3.14	.49	.71

Note. [a]N = 368. [b]N = 425. [c]N = 413.

Norms

In reports provided to students, scores are conveyed using a four-point rating system that mimics a gas gauge. Scores fall into one of four possible categories: Developing (scores in approximately the 16th percentile or lower), Approaching (scores in approximately the 17th–50th percentiles), Demonstrating (scores in approximately the 51st–84th percentiles), and Mastering (scores at or above the 85th percentile). The field study sample of 1,206 students is used as the norming sample to determine each cut score. The raw scores that correspond to each range are as follows:

- Developing: raw scores at or below −.9158
- Approaching: raw scores between −.91581 and .1403
- Demonstrating: raw scores between .14031 and .8945
- Mastering: raw scores at or above .89541

References

Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., & De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological Assessment, 31*(4), 1–14.

ACT. (2015). *Beyond academics: A holistic framework for enhancing education and workplace success.* Iowa City, IA: ACT.

ACT. (2016). *Development and validation of ACT Engage®: Technical manual.* Iowa City, IA: ACT.

ACT. (2021). *Mosaic^{TM} by ACT®: Social Emotional Learning Assessment technical manual.* Iowa City, IA: ACT.

Allen, J., Way, J. D., & Casillas, A. (2019). Relating school context to measures of psychosocial factors for students in grades 6 through 9. *Personality and Individual Differences, 136*(1), 96–106.

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences, 34*(4), 645–664.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? International *Journal of Selection and Assessment, 9*(1–2), 9–30.

Brackett, M. A., Elbertson, N. A., & Rivers, S. E. (2015). Applying theory to the development of approaches to SEL. In J. A. Durlak, C. E. Domotrovich, R. P. Weissberg, & T. P. Gullotta (Eds.) *Handbook of social and emotional learning: Research and practice.* pp. 20–32. New York, NY: The Guilford Press.

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18*(1), 36–52.

Campion, M. C., Ployhart, R. E., & MacKenzie, W. I., Jr. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance, 27*(4), 283–310.

Casillas, A., Way, J., & Burrus, J. (2015). Behavioral skills. In W. Camara, R. O'Connor, K. Mattern, & M. A. Hanson (Eds.), *Beyond academics: A holistic framework for enhancing education and workplace success* (pp. 25–38). Iowa City, IA: ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2015-4.pdf.

Chernyshenko, O., Kankaraš, M., & Drasgow, F. (2018*). Social and emotional skills for student success and wellbeing: Conceptual framework for the OECD study*

on social and emotional skills (OECD Education Working Papers, No. 173). Paris, France: Organisation for Economic Co-operation and Development. https://dx.doi.org/10.1787/db1d8e59-en.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112(*1), 155.

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review, 25,* 56–72.

Devine, J., & Cohen, J. (2007). *Making your school safe: Strategies to protect children and promote learning.* New York, NY: Teachers College Press.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta analysis of school based universal interventions. *Child Development, 82*(1), 405–432.

Gregory, A., & Cornell, D. (2009). "Tolerating" adolescent needs: Moving beyond zero tolerance policies in high school. *Theory into Practice, 48*(2), 106–113.

Halverson, C. F., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlpoulos, V., Besevegis, E., & Wen, L. (2003). Personality structure as derived from parental ratings of free descriptions of children: The Inventory of Child Individual Differences. *Journal of Personality, 71*(6), 995–1026.

Heaven, P. C. L., Mak, A., Barry, J., & Ciarrochi, J. (2002). Personality and family influences on adolescent attitudes to school and self-related academic performance. *Personality and Individual Differences, 32,* 453–462.

Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of Situational Judgment Tests: Faking, coaching, and retesting issues. In J. Weekley & R. Ployhart (Eds.), *Situational judgment tests* (pp. 205–323). New York, NY: Psychology Press.

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology, 1*(3), 272–290.

Hudson, N. W., Briley, D. A., Chopik, W. J., & Derringer, J. (2019). You have to follow through: Attaining behavioral change goals predicts volitional personality change. *Journal of Personality and Social Psychology, 117(*4), 839.

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). *Personality, motivation, and college readiness: A prospectus for assessment and development* (ETS RR-14-06). Princeton, NJ: Educational Testing Service.

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality inventories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York, NY: Guilford Press.

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*(3), 547–561.

Mervielde, I., & De Fruyt, F. (1999). Construction of the hierarchical personality inventory for children (HiPIC). In I. Mervielde, I. Deary, F. de Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (proceedings of the eighth European Conference on Personality Psychology). Tilburg University Press, Tilburg.

Murano, D., Lipnevich, A. A., Walton, K. E., Burrus, J., Way, J. D., & Anguiano-Carrasco, C. (2021). Measuring social and emotional skills in elementary students: Development of self-report Likert, situational judgment test, and forced choice items. *Personality and Individual Differences, 169*, 110012.

OECD. (2012). P*ISA 2009 technical report.* Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264167872-en.

Osher, D., & Berg, J. (2017). *School climate and social and emotional learning: The integration of two approaches.* University Park, PA: Edna Bennet Pierce Prevention Research Center, Pennsylvania State University.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401–421.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338.

Primi, R., Santos, D., John, O. P., & De Fruyt, F. (2016). Development of an inventory assessing social and emotional skills in Brazilian youth. *European Journal of Psychological Assessment, 32(*1), 5–16.

Rimm-Kaufman, S. E., & Hulleman, C. S. (2015). SEL in elementary school settings: Identifying mechanisms that matter. In J. A. Durlak, C. E. Domotrovich, R. P. Weissberg, & T. P. Gullotta (Eds.) *Handbook of social and emotional learning: Research and practice* (pp. 151–166). New York, NY: The Guilford Press.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313–345.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*(2), 117–141.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1), 1–25.

Roberts, R. D., Martin, J. E., & Olaru, G. (2015). *A Rosetta stone for noncognitive skills: Understanding, assessing, and enhancing noncognitive skills in primary and secondary education.* New York, NY: Asia Society, Professional Examination Service. Retrieved from: https://asiasociety.org/files/A_Rosetta_Stone_for_Noncognitive_Skills.pdf.

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813.

Schmitt, D., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*(2), 173–212.

Skinner, E., & Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571–581.

Soto, C. J., Napolitano, C. M., & Roberts, B. W. (2021). Taking skills seriously: Toward an integrative model and agenda for social, emotional, and behavioral skills. *Current Directions in Psychological Science, 30*(1), 26–33.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. Psychological Bulletin, 134(1), 138–161.

Tackett, J. L., Slobodskaya, H. R., Mar, R. A., Deal, J., Halverson, C. F., Jr., Baker, S. R., ... & Besevegis, E. (2012). The hierarchical structure of childhood personality in five countries: Continuity from early childhood to early adolescence. *Journal of Personality, 80*(4), 847–879.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171.

Walton, K. E., Burrus, J., Anguiano-Carrasco, C., Way, J. D., & Murano, D. (2019). *Aligning ACT Tessera to the Collaborative for Academic, Social, and Emotional Learning (CASEL) framework.* Iowa City, IA: ACT.

Walton, K. E., Murano, D., Burrus, J., & Casillas, A. (2021). *Multi-method support for using the Big Five framework to organize social and emotional skills.* Assessment. Advance online publication.

Ziegler, M., MacCann, C., & Roberts, R. D. (2012). *New perspectives on faking in personality assessment.* Chicago, IL: Oxford University Press.