Mosaic™ by ACT®: Social Emotional Learning Assessment

Spring 2021

Version 2021.1

Mosaic™ by ACT®
Social Emotional Learning

www.act.org

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.

Version	Date	Type	Description
2021.1	Spring 2021	Interim update	Rebranding and minor updates.
2020.2	Fall 2020	Interim update	Branding and minor text updates.
2020.1	Spring 2020	Annual update	General updates.

ACT, Inc. 2021

© by ACT, Inc. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.https://creativecommons.org/licenses/by-nc/4.0/

ACT.org/research

Table of Contents

Chapter 1
Introduction . . . 1.1

Chapter 2
Mosaic by ACT: Social Emotional Learning Assessment Background and Social Emotional Skills Assessed . . . 2.1

What are Social Emotional Skills? . . . 2.1
What is School Climate? . . . 2.2
Assessment Framework: Five Factor Model of Personality . . . 2.2
Five Factor Model Overview . . . 2.2
Five Factor Model Traits' Associations with Academic Outcomes . . . 2.4
Five Factor Model Traits' Development over the Lifespan . . . 2.5
Assessment Skills and their Alignment to the Five Factor Model . . . 2.7
Climate Dimensions . . . 2.7

Chapter 3
Multi-Trait Multi-Method Assessment . . . 3.1

Likert Items . . . 3.1
Forced Choice Items . . . 3.2
Situational Judgment Tests . . . 3.4
Unified Score . . . 3.5
Collaborative for Academic, Social, and Emotional Learning (CASEL) Core Competency Composite Scores . . . 3.5

Chapter 4
Initial Item Development . . . 4.1

Initial Item Development . . . 4.1
Item Parameter Estimation . . . 4.1
Final Item Set . . . 4.2

Chapter 5
Reliability and Validity Evidence . . . 5.1

Procedure . . . 5.1
Middle School Assessment . . . 5.2
Method . . . 5.2
Results . . . 5.2
Subgroup Differences . . . 5.7
Norms . . . 5.10
High School Assessment . . . 5.10
Method . . . 5.10
Results . . . 5.11
Subgroup Differences . . . 5.16
Norms . . . 5.19

References . . . R.1

List of Tables

Table 1. Meta-analytic Correlations Between the Big Five and Cognitive Ability and Grade Point Average (Poropat, 2009) 2.4

Table 2. Alignment of Assessment Skills to the Five Factor Model 2.7

Table 3. Alignment of Assessment Skills to the CASEL Core Competencies 3.6

Table 4. Correlations Between Assessment Skills and School Climate Dimensions 5.4

Table 5. Correlations Between Middle School Social Emotional Skills and School Climate Dimensions and GPA. 5.5

Table 6. Associations Between Middle School Social emotional Skills and School Climate Dimensions and Discipline. 5.7

Table 7. Gender Differences on Middle School Assessment Scales. 5.8

Table 8. Free/Reduced Lunch Status Differences on Middle School Assessment Scales 5.9

Table 9. Race/Ethnicity Differences on Middle School Assessment Scales 5.10

Table 10. Correlations Between High School Social Emotional Skills and School Climate Dimensions 5.12

Table 11. Correlations Between High School Social Emotional Skills and School Climate Dimensions and GPA. 5.12

Table 12. Associations between High School Scales and Climate and Absenteeism. 5.14

Table 13. Associations Between High School Social Emotional Skills and School Climate Dimensions and Discipline. 5.16

Table 14. Gender Differences on the High School Assessment and ACT Aspire 5.17

Table 15. Free/Reduced Lunch Status Differences on the High School Assessment. 5.18

Table 16. Race/Ethnicity Differences on the High School Assessment and ACT Aspire 5.19

List of Figures

Figure 1. Meta-Analytic Evidence Showing Personality Changes over the Lifespan 2.5

Figure 2. Example Likert Item 3.2

Figure 4. Example Situational Judgment Test Items 3.4

Figure 5. Unified Scoring Model 3.7

Figure 6. Average Self-Reported GPA by Quartiles of Middle School Social Emotional Skills and School Climate Scores 5.5

Figure 7. Average Self-Reported GPA by Quartiles of High School Social Emotional Skills and School Climate Scores. 5.13

Figure 8. Associations between High School Social Emotional Skills and Absenteeism 5.15

Figure 9. Associations between High School Climate Dimensions and Absenteeism 5.15

Chapter 1

Introduction

Mosaic™ by ACT®: Social Emotional Learning Assessment is a comprehensive assessment system designed to measure five social emotional skills and two dimensions of school climate. The online assessment provides students, parents, teachers, and schools with a holistic picture of students' strengths and areas for improvement and their perception of school safety and quality of relationships with school personnel. The assessment system includes student- and school-level reports that provide feedback on each skill. Mosaic by ACT: Social Emotional Learning also has a curriculum component designed to teach social emotional skills. The assessment can be used in conjunction with this curriculum as part of a comprehensive set of social emotional learning tools.

This technical manual details the empirical basis for the assessment framework, the constructs assessed, the various item types used, and the procedure used to incorporate the item types into a unified score (for each of the five skill areas). In addition, evidence of reliability and validity is provided, as in normative data.

Chapter 2

Mosaic by ACT: Social Emotional Learning Assessment Background and Social Emotional Skills Assessed

What are Social Emotional Skills?

There is growing consensus in the realm of public policy and in research in education, psychology, and economics that a number of factors outside of cognitive ability may be nearly, or just as important for educational and workplace success (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). A few examples of these factors include working hard, getting along with others, and persevering through challenges. Given that most of these factors often demonstrate low zero-order correlations with cognitive ability, they are often referred to as noncognitive factors (see Kyllonen, Lipnevich, Burrus, & Roberts, 2014, for a review). Several alternative terms exist, including character strengths, social emotional skills, personal skills, personal qualities, and psychosocial skills (Kyllonen et al., 2014).

In the interest of clarity, we will use the term social emotional skills, as this term resonates clearly with educators and is common in policy settings. Social emotional skills can be defined as: “individual capacities that (a) are manifested in consistent patterns of thoughts, feelings, and behaviours, (b) can be developed through formal and informal learning experiences, and (c) influence important socioeconomic outcomes throughout the individual’s life” (Organisation for Economic Co-operation and Development [OECD], 2015, p. 34).

Educators nearly universally believe in the value of social emotional skills, and this belief is reflected in their presence in K-12 (e.g., Stemler & Bebell, 2012) and university mission statements (e.g., Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). Policymakers have also recognized the value of social emotional skills. The most recent reauthorization of the federal Elementary and Secondary Education Act of 1965, the Every Student Succeeds Act (ESSA), provides states and districts with increased flexibility on the use of federal funds and accountability measures, both of which can be used to support the development of social emotional skills in schools (Grant et al., 2017). This growing consensus on the value of social emotional skills is supported by decades of research. To echo the OECD definition, social emotional skills predict a variety of important outcomes. These outcomes include, but are not limited to:

- Academic performance such as grades (Poropat, 2009)
- Academic retention (Robbins et al., 2004)
- Behavioral problems (Ge & Conger, 1999)
- Happiness (Diener & Lucas, 1999)
- Health (Bogg & Roberts, 2004) and longevity (Roberts et al., 2007)
- Job performance (Barrick, Mount, & Judge, 2001)
- Job satisfaction (Judge, Heller, & Mount, 2002)
- Marital satisfaction (Watson, Hubbard, & Wiese, 2000)
- Peer relationships (Jensen-Campbell et al., 2002)

Furthermore, in line with the OECD definition, there is evidence that social emotional skills change over time (Roberts, Walton, & Viechtbauer, 2006) and may be improved through school-based programs (e.g., Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

What is School Climate?

In addition to measuring individual differences in students' skills, it is important to measure students' perception of the larger context within which they are receiving their education. One approach to this is to assess student perceptions of school climate. School climate can include several factors, such as emotional and physical safety, connectedness, respect, engagement, and challenge (Berg, Osher, Moroney, & Yoder, 2017). Similarly, a review of 206 papers found that school climate tends to focus on five dimensions: safety, relationships, teaching and learning, institutional environment, and the school improvement process (Thapa, Cohen, Guffey, & Higgins-D'Alessando, 2015). School climate is related to a number of outcomes in schools. To name a few, positive school climate is related to higher student self-esteem (Hoge, Smit, & Hanson, 1990), better psychological well-being (Ruus et al., 2007), decreased absenteeism (e.g., ACT, 2016), and lower rates of student suspension (e.g., Lee, Cornell, Gregory, & Fan, 2011).

Assessment Framework: Five Factor Model of Personality

Five Factor Model Overview

The assessment uses the Five Factor Model of personality (FFM; Digman, 1990) as its organizing framework. The FFM was chosen because it is an evidence-based framework that helps K-12 policymakers and educators make sense of the plethora of social emotional skill labels that are used across the field. It effectively organizes the many skills that are critical for student success. In addition, there is a large body of psychological research conducted in the past few decades supporting the framework and establishing the many critical education and life outcomes associated with the five personality factors. The origins of the FFM are discussed below.

Guided by the lexical hypothesis, which assumes that important individual differences will become encoded into language as single terms (Goldberg, 1993), Allport and Odbert (1936) searched Webster's 1925 New International Dictionary for English words describing human characteristics. In total, roughly 18,000 English words were selected, with 4,500 being classified as descriptions of stable personal traits. Cattell (1943) applied factor analytic procedures to reduce the massive list of traits by analyzing the underlying patterns among them. Subsequent efforts to reduce the massive list resulted in five replicable factors (Fiske, 1949; Norman, 1963; Tupes & Christal, 1961). Based on the item content of each factor, they are most commonly labeled extraversion, agreeableness, conscientiousness, emotional stability (often referred to by its opposite pole, neuroticism), and openness to experience, and are often referred to as the Big Five (see de Raad & Mlačić, 2015, for a complete history).

Extraversion describes a person's likelihood to engage in social interaction, be active and assertive, and experience positive emotionality. Highly extraverted people tend to enjoy leadership positions, feel comfortable expressing unpopular opinions, and express positive emotions. Agreeableness is associated with a prosocial and communal orientation towards others. A very agreeable person can be described as friendly, helpful, and empathic. Conscientiousness primarily describes the tendency to control impulses in an effort to achieve one's goals. A highly conscientious individual can be described as organized, diligent, and rule following. Emotional stability describes a person's capability to cope with stressful situations and emotions. Emotionally stable individuals are able to successfully cope with stressful situations and do not experience negative emotions, such as depression, stress, or anxiety, to an overwhelming degree. Openness to experience is somewhat related to cognitive ability (Ackerman & Heggestad, 1997) and can best be described as a person's interest in and acceptance of unfamiliar cultures, ideas, values, artistic endeavors, and even feelings. Highly open individuals have original, creative, and complex thoughts. Full descriptions and behavioral examples of the Big Five are offered by John, Naumann, and Soto (2008).

Although the Big Five were first discovered in the English language, studies in other languages or cultures, either involving the full psycholexical approach (for an overview see de Raad, 2000), or applying translations of established FFM inventories, have resulted in the same five factors (see, e.g., McCrae et al., 2005; Schmitt, Allik, McCrae, & Benet-Martinez, 2007). As one example, McCrae and colleagues (2005) documented the replication of the FFM in nearly 50 cultures on six continents. There is overwhelming evidence supporting the FFM's universality in the vast majority of countries, cultures, and languages around the world.

In addition to its empirical support, the popularity of the FFM has grown because it optimizes bandwidth and fidelity (Cronbach & Gleser, 1965). That is, it allows for the summary of a large amount of information while simultaneously allowing for some nuanced individual difference description (Soto & John, 2014). The Big Five traits lie at what some have called the "foundational level" (Soto & Tackett, 2015) of what is a hierarchical structure of personality traits (Markon, 2009; Markon, Krueger, & Watson, 2005). There are fewer than five constructs at the broad, higher level of the hierarchy (e.g., alpha and beta) and more than five at the narrow, lower level of the hierarchy. The constructs at the lower level, which are typically referred to as facets, allow for more nuanced individual difference description and greater predictive validity than broad traits (Paunonen & Ashton, 2001). However, with a large number of facets (e.g., 30 in the Revised NEO Personality Inventory; Costa & McCrae, 1995), personality description becomes more cumbersome and personality assessment requires more resources.

Given that the Big Five are broad and multifaceted, this model can be used to integrate the plethora of social emotional skill terms discussed in the literature or assessed in practice. The Big Five can be considered as something of a "Rosetta Stone" for understanding social emotional skills (Martin et al., 2019). The Rosetta Stone allowed archaeologists to understand how languages related to one another and how words in different languages had the same underlying meaning. Using the Big Five, we can take constructs expressed as time management in one framework, grit in another, and responsibility in yet a third, and understand their connectedness by seeing them all as manifestations or facets of conscientiousness, at least in significant measure. That is, it is likely that most or all social emotional skills can be mapped to the Big Five.

Five Factor Model Traits' Associations with Academic Outcomes

There is a vast body of psychological research linking the Big Five with many critical outcomes, including those in the academic domain. Table 1 provides a summary of meta-analytic data on the Big Five-academic performance link compiled with an accumulated sample size of over 70,000 students (Poropat, 2009). Specifically, it displays the correlations between the Big Five and grade point average (GPA) and the correlation between cognitive ability and GPA. The meta-analytic evidence is compelling; these traits are key for academic success. During the primary school years, cognitive ability's impact on academic performance exceeds that of any Big Five trait, but by secondary education, conscientiousness is nearly as important for academic performance as cognitive ability. However, it receives less attention in large-scale assessments with policy impact. Moreover, little attention is paid to how these types of skills might be enhanced during a student's school career via formal instruction and universal or targeted interventions.

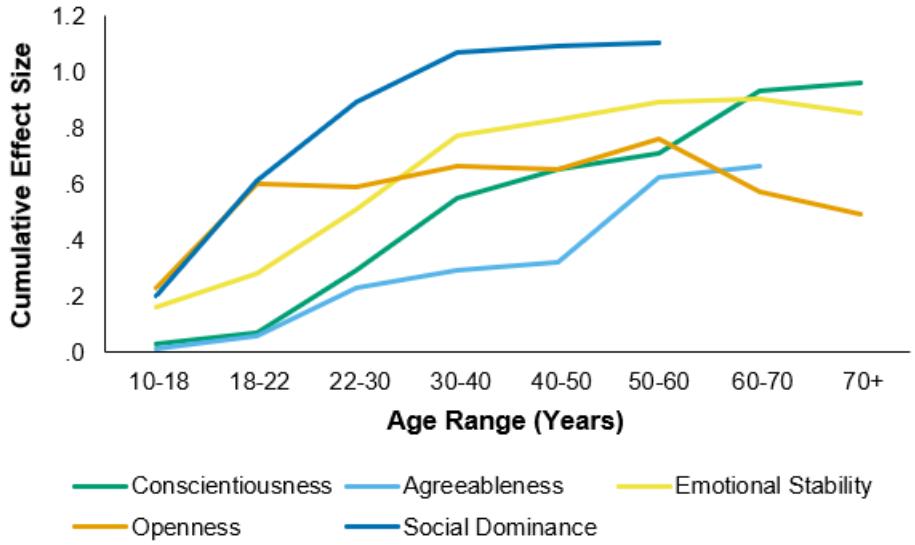
Table 1. Meta-analytic Correlations Between the Big Five and Cognitive Ability and Grade Point Average (Poropat, 2009)

	Education Level		
	Primary	Secondary	Tertiary
Conscientiousness	.28	.21	.23
Agreeableness	.30	.05	.06
Emotional Stability	.20	.01	-.01
Openness to Experience	.24	.12	.07
Extraversion	.18	-.03	-.01
Cognitive Ability	.58	.24	.23

Five Factor Model Traits' Development over the Lifespan

The last point made in the preceding section is particularly pertinent. The fact that there is not tremendous effort made to enhance students' social emotional skills is likely due to the long-standing notion personality is "set like plaster" (Costa & McCrae, 1994; James, 1981). That is, until recently, many held the belief that personality traits or social emotional skills are immutable, particularly as one ages. People's tendency to hold on to this view may stem from the fact that there is some trait consistency over time. For example, Roberts and DelVecchio (2000) examined 152 longitudinal studies to show that the rank-order consistency of personality was fairly moderate with estimated test-retest correlations of .31 in early childhood (0-2.9 years) and generally increasing over time, reaching .74 in the 50-59 decade when consistency estimates plateau. This level of consistency has been misinterpreted as indicating that personality does not change over time. However, over time, individuals can maintain their relative standing within a group while the entire group can shift a nontrivial degree. That is, high rank-order consistency does not imply that there is no mean-level change. In their meta-analysis of mean-level personality change, Roberts et al. (2006) found that individuals become more socially dominant (a facet of extraversion), conscientious, agreeable, and emotionally stable throughout the lifespan, particularly in adolescence and early adulthood, and the effects were not slight; change over the lifespan reached a full standard deviation (see Figure 1).

Figure 1. Meta-Analytic Evidence Showing Personality Changes over the Lifespan



Note: Roberts et al. (2006) examined mean-level change of two components of extraversion with different developmental patterns, namely social dominance and social vitality. Here we present the findings for social dominance.

Research in this field has led to the conclusion that cognitive ability may not be changed easily (Kyllonen, Roberts, & Stankov, 2008), but personality traits can be, and in fairly brief and sometimes innocuous ways (Dweck, 2012). Roberts and colleagues carried out a meta-analysis showing that interventions (primarily clinical interventions) have the capacity to alter personality traits to a significant degree even when the intervention has a duration of just eight weeks (Roberts et al., 2017). Outside the

realm of clinical interventions, research suggests that personality trait change can occur as the result of something fairly simple – specifically, completing “challenges” (Hudson, Briley, Chopik, & Derringer, 2019). The challenges Hudson and colleagues used were specific, concrete actions that were small and reasonable for someone to complete, such as: “ask a friend to go for coffee” or “read a news story about a foreign country.” Completing two extraversion challenges per week, for example, resulted in an average increase of .17 standard deviations over the course of a single semester.

Another recent meta-analysis sheds light on the effectiveness of social emotional learning interventions. Summarizing the results of over 75 studies, including studies of afterschool programs where social emotional skills are inculcated, Durlak, Weissberg, and Pachan (2010) noted that these informal learning programs had an overall positive and statistically significant impact on participants. These changes did not occur in all domains, but rather in three main areas: feelings and attitudes, indicators of behavioral adjustment, and school performance. In particular, there were significant increases in youths’ self-perceptions, bonding to school, positive social behaviors, school grades, and achievement test scores. There were also significant reductions in problem-related behaviors. In addition, certain programs that used a protocol focused on sequenced, active, focused, and explicit (i.e., SAFE) programming were associated with practical gains in participants’ test scores of 12 percentile points (compared to control groups’ gains), a result that is similar to or better than those obtained by many other evidence-based interventions for school-aged populations. Durlak et al. (2010) concluded that current findings for afterschool programs “merit support and recognition as an important community setting for promoting youths’ personal and social well-being and adjustment” (p. 302), a finding which has been reported in several meta-analyses (Mahoney, Durlak, & Weissberg, 2018). Moreover, it should be noted that the economic benefits of investing in social emotional learning programs outweigh the costs 11:1; that is, for every $1 spent on such programs, there will be an $11 gain (Belfield et al., 2015).

This body of research supports the potential for social emotional learning interventions delivered in educational contexts. Although the effectiveness of social emotional learning programs was once dubious, there now appears to be a solid evidentiary base demonstrating that they are not only plausible but are also credible through secondary and even postsecondary education. Given the high valuation of social emotional skills by educators, the effectiveness of high quality social emotional learning programs, and the relative cost effectiveness of such programs, social emotional learning should play a more pivotal role in educational policy and practice than previously realized.

Assessment Skills and their Alignment to the Five Factor Model

The constructs assessed by the assessment can be aligned with FFM constructs on a one-to-one basis. This alignment is provided in Table 2.

Table 2. Alignment of Assessment Skills to the Five Factor Model

Assessment Skill	Skill Definition	Big Five Factor
Sustaining Effort	The extent to which a student's actions demonstrate persistence, goal striving, reliability, dependability, and attention to detail at school.	Conscientiousness
Getting Along with Others	The extent to which a student's actions demonstrate collaboration, empathy, helpfulness, trust, and trustworthiness.	Agreeableness
Maintaining Composure	The extent to which a student's actions demonstrate stress management, emotional regulation, a positive response to setbacks, and poise.	Emotional stability
Keeping an Open Mind	The extent to which a student's actions demonstrate creativity, inquisitiveness, flexibility, open-mindedness, and embracing diversity.	Openness to experience
Social Connection	The extent to which a student's actions demonstrate assertiveness, influence, optimism, and enthusiasm.	Extraversion

This alignment was conducted rationally by comparing the assessment skills' and the FFM factors' definitions. Moreover, this alignment is supported by research demonstrating significant correlations between the assessment skills and the FFM factors. Some examples include:

- Grit (related to Sustaining Effort) is correlated with conscientiousness (Duckworth & Quinn, 2009)
- Interpersonal conflict (related to Getting Along with Others) is correlated with agreeableness (Jensen-Campbell & Graziano, 2001)
- Coping (related to Maintaining Composure) is correlated with emotional stability (MacCann, Lipnevich, Burrus, & Roberts, 2012)
- Curiosity (related to Keeping an Open Mind) is correlated with openness to experience (Komarraju, Karau, & Schmek, 2009)
- Social Connection is correlated with extraversion (Judge, Bono, Ilies, & Gerhardt, 2002)

Climate Dimensions

The assessment measures the Relationships and Safety dimensions of school climate. 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 level of absenteeism and lower levels of academic achievement (ACT, 2016). Importantly, these dimensions are also related to social 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 emotional skills have a bidirectional relationship, with the one supporting the other (Osher & Berg, 2017).

Chapter 3

Multi-Trait Multi-Method Assessment

Mosaic by ACT: Social Emotional Learning Assessment measures the five social emotional skills with three methods: Likert items, forced choice items, and situational judgment tests (SJT). Every measurement method is subject to its own biases or weaknesses, and the assessment employs multiple methods to minimize the effects of these biases or weaknesses. This is known as a multi-trait multi-method (MTMM) design. According to Kenny and Kashy (1992), "The underlying view of measurement in the MTMM analysis is that to measure a theoretical construct, different measures, each with its own bias, are selected. Bias that is due to method effects is reduced through a triangulation process" (p. 170).

The benefit of the MTMM design is intuitive. For example, Likert items might be easily faked (see below), whereas SJTs might be somewhat more difficult to fake but might have problems of their own, such as some students having difficulty mentally projecting themselves into hypothetical situations. If only one of these methods is employed in an assessment system, then the bias associated with that method can seriously compromise the assessment's validity. However, the use of multiple methods minimizes this problem. Each assessment method is briefly described below.

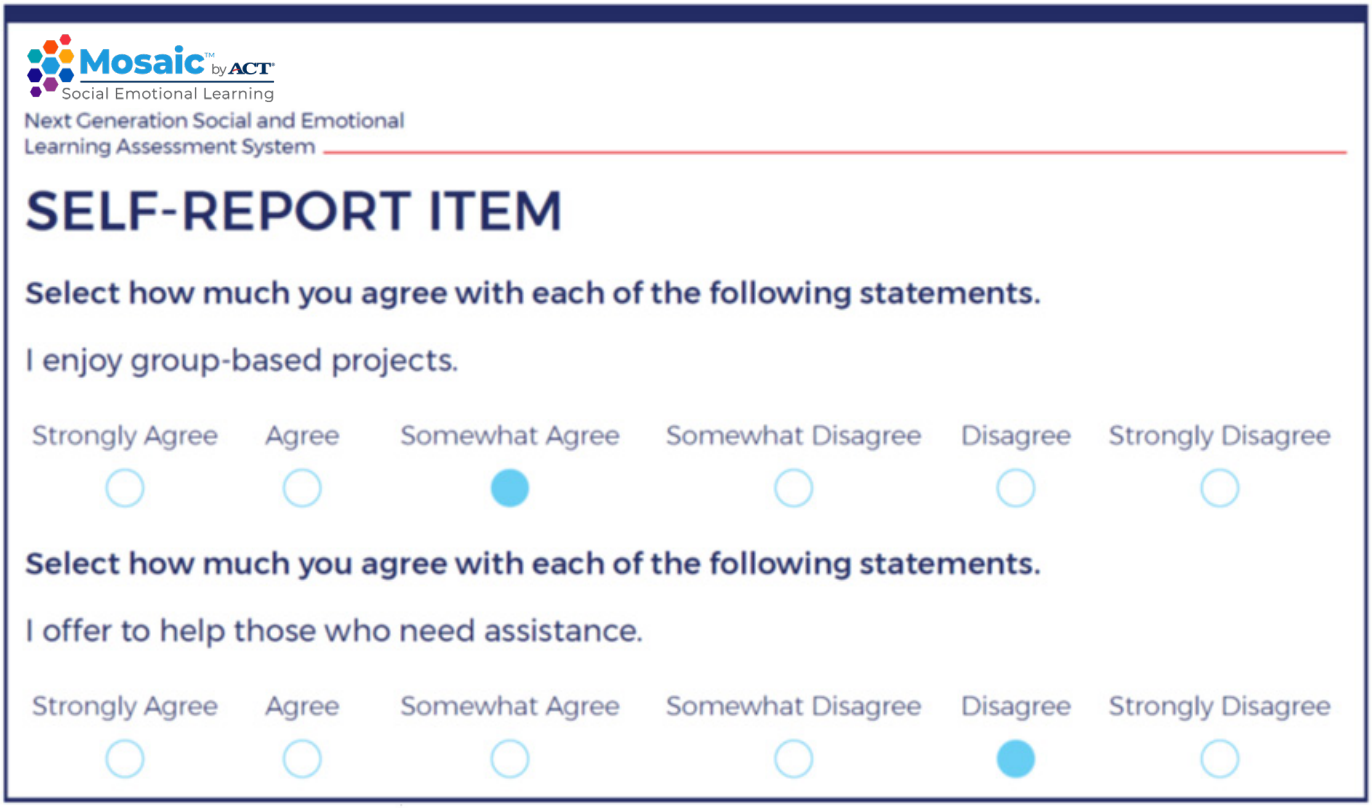
Likert Items

Likert items have been used in social emotional learning research and practice for decades and are known to allow one to efficiently gather a lot of information in a brief period of time. Individuals are asked to indicate their level of agreement with a number of statements (e.g., "I work hard at school"; see Figure 3 for an example).

This type of assessment is preferred in environments when there are no stakes for the self-assessor and faking is not expected (Lipnevich, MacCann, & Roberts, 2013). However, respondents may have various motives for faking their responses, such as to avoid having to attend training programs or to appear more attractive to a prospective school admissions officer, university system, or employer (e.g., Zickar, Gibby, & Robie, 2004). Furthermore, Likert items might be particularly susceptible to reference effects. That is, often people answer such items by asking the question, "compared to whom?" As a consequence, it could be the case that students from very high achieving schools, for example, might rate themselves lower on their social emotional skills than students from low-achieving schools simply because they are using a different

reference group and not because they are truly lower on these skills. This is often called the Big-Fish Little-Pond Effect (Marsh & Hau, 2003).

Figure 2. Example Likert Item



Forced Choice Items

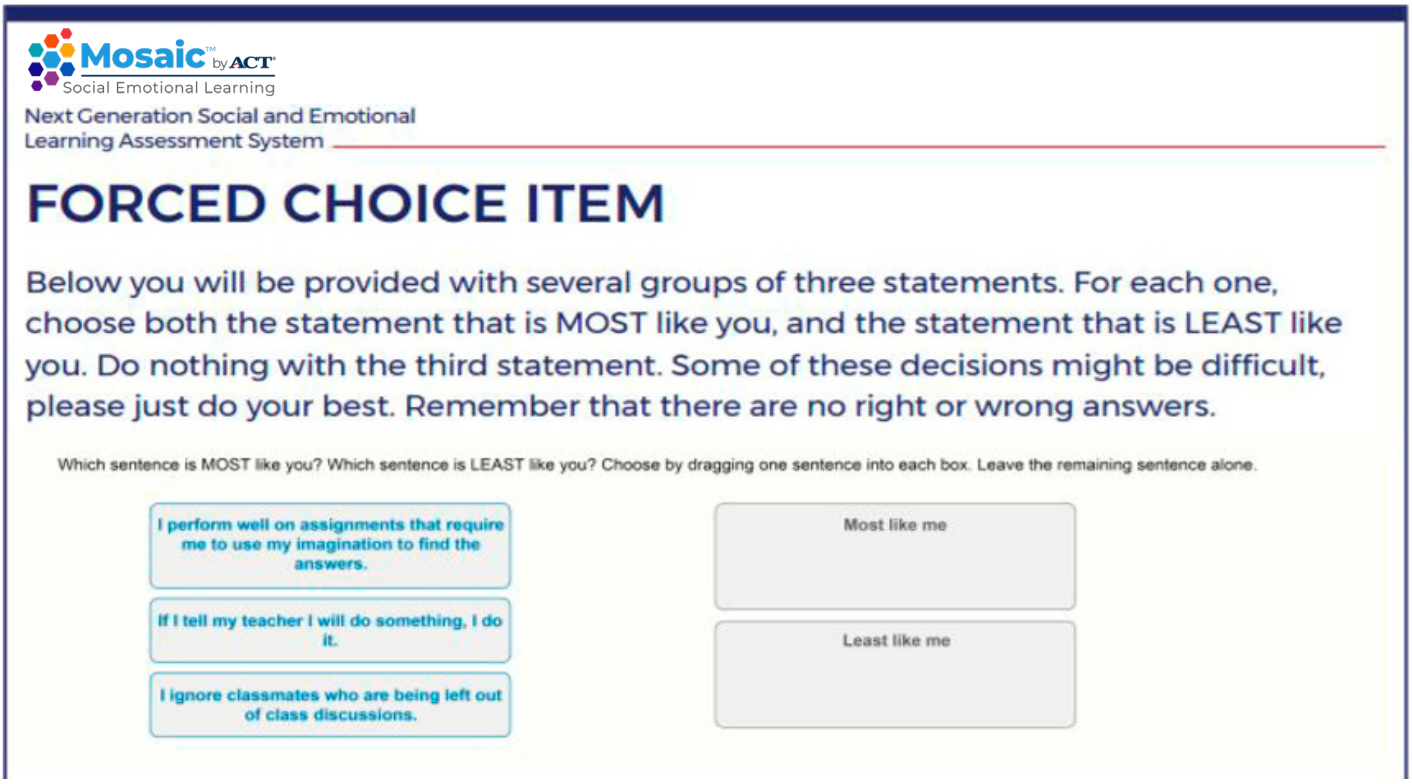
In forced choice (FC) items, statements (or sometimes adjectives) are grouped in blocks, and respondents are instructed to make selections within each block regarding which statements describe him or her best. There are several variations of FC methodology (Hontangas et al., 2015). One variation is the number of items included per block. For example, in pair comparisons, the respondent must choose between two items (e.g., Which is more like you: “I enjoy leading class discussions” or “I work hard in school to achieve my goals”?). Often, three or four items appear per block. FC inventories can also vary according to instructions given to the respondents; they may be asked to either fully or partially rank order the items from “least like me” to “most like me,” or they may be asked to select just one item that is characteristic of them. Finally, there can be either multidimensional or unidimensional forced choice assessments. In a unidimensional FC block, all statements measure the same latent construct (e.g., “I prefer to work in a group” vs. “I prefer to work alone”). In multidimensional FC assessments, respondents are presented with two or more items, each of which is an indicator of a different latent trait. Any permutation of these three assessment characteristics can be selected based on one’s goals and targeted population characteristics. An example of a partially ranked multidimensional FC triad is

presented in Figure 4. Respondents are instructed to select the statement that is “most like me” and the one that is “least like me.”

There is compelling evidence to suggest that FC items cannot be faked as easily as Likert items (Christiansen, Burns, & Montgomery, 2005; Jackson, Wroblewski, & Ashton, 2000; Walton et al., 2019). There is also some evidence that they have stronger relationships with performance outcomes (Drasgow et al., 2012). Furthermore, given that no scales are used in these items, FC tests eliminate scale response effects. Reference bias should also be minimized with FC tests because respondents conduct an internal (self vs. self) rather than an external (self vs. other) comparison when responding to the items.

Until recently, this methodology had one notable downside; only ipsative scores could be produced, which do not allow for between-person comparisons and have several psychometric limitations (Meade, 2004). Recent advances, however, allow for normative scores to be generated through item response theory (IRT) modeling (Maydeu-Olivares & Brown, 2010; Wang, Lee, Joo, Stark, & Louden, 2016; Wang, Qiu, Chen, Ro, & Jin, 2017). Several such models exist (McCloy, Heggestad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005), and one of the more widely used models is a Thurstonian IRT model (Brown & Maydeu-Olivares, 2011), which entails modeling binary outcomes, resulting from pairwise comparisons within the same block. It is a multidimensional normal ogive model with local dependencies due to statements appearing in the multiple pairs (utilities) associated with each item having constrained (equal) parameter estimates. Using this methodology, one can compute item loadings, item thresholds, and factor scores, which translate to item discrimination, item difficulty, and person parameters (trait scores), respectively, in traditional IRT terminology.

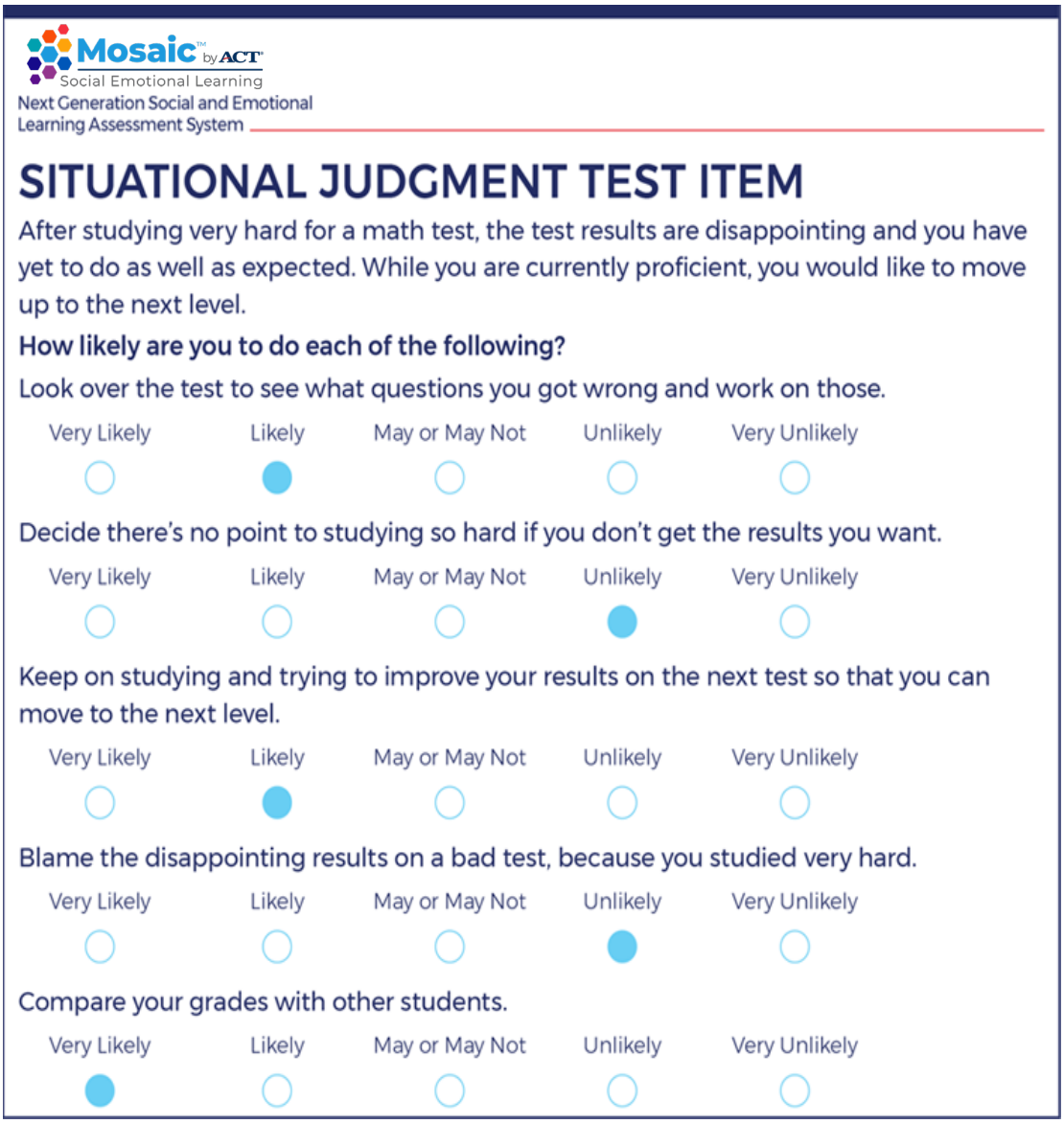
Figure 3. Example Forced Choice Triad



Situational Judgment Tests

A situational judgment test (SJT) is one in which participants are asked how best to, or how they might typically, deal with some situation (see Figure 5 for an example). Situations can be described in writing or can be audio/videotaped, and response types can include multiple choice, constructed response, and ratings (McDaniel, Morgesen, Finnegan, Campion, & Braverman, 2001). The methodology has been used to assess many relevant attributes of individuals, including leadership, teamwork, and conscientiousness (Kyllonen & Lee, 2005; MacCann & Roberts, 2008; Oswald et al., 2004; Wang, MacCann, Zhuang, Liu, & Roberts, 2009). SJTs have been shown to predict many criteria, such as academic success (Lievens & Sackett, 2012; Oswald et al., 2004) and managerial performance (Howard & Choi, 2000).

Figure 4. Example Situational Judgment Test Items



SJTs have several additional advantages over traditional Likert scales. First, SJTs may be developed to reflect both general and more subtle and complex judgment processes than what is possible with conventional Likert scales. Second, SJTs appear to be associated with less adverse impact on ethnic minorities, which may be of relevance for mitigating subgroup differences in any population under consideration (Schmitt et al., 2009). Third, SJTs can be repurposed as formative assessments so as to provide a student with feedback on his or her competencies in the domain of interest. Fourth, SJTs have the advantage of face validity; that is, the situations presented to students "look and feel" like situations that would be encountered in real life. Fifth, there is evidence suggesting SJTs are less prone to faking than Likert items (Hooper, Cullen & Sackett, 2006). Finally, students report they are engaging and worth completing (Lipnevich et al., 2013), which better supports multiple administrations and retains student "buy-in" to the ongoing process of social emotional learning assessment. SJTs are not without their own limitations, however. For example, they are often multidimensional (McDaniel & Whetzel, 2007) and may have a fairly high reading load, which can be taxing for the respondent.

Unified Score

The three item types are combined to yield a single, unified score per social emotional skill. To obtain the unified score, item-level data are entered into a confirmatory factor model. Each Likert item loads on the target factor, each SJT behavioral response loads on the target factor as well as a factor representing the SJT, and each binary pairwise comparison produced by the FC triads loads on the two target factors, per the Thurstonian IRT model (see Figure 5). From this model, parameters are estimated for each skill then transformed into multidimensional item response model parameters. Although relatively new, there is evidence of the criterion validity of the unified scores, and the unified scores have greater predictive validity than single item types (Anguiano-Carrasco, Walton, Murano, Burrus, & Way, 2018).

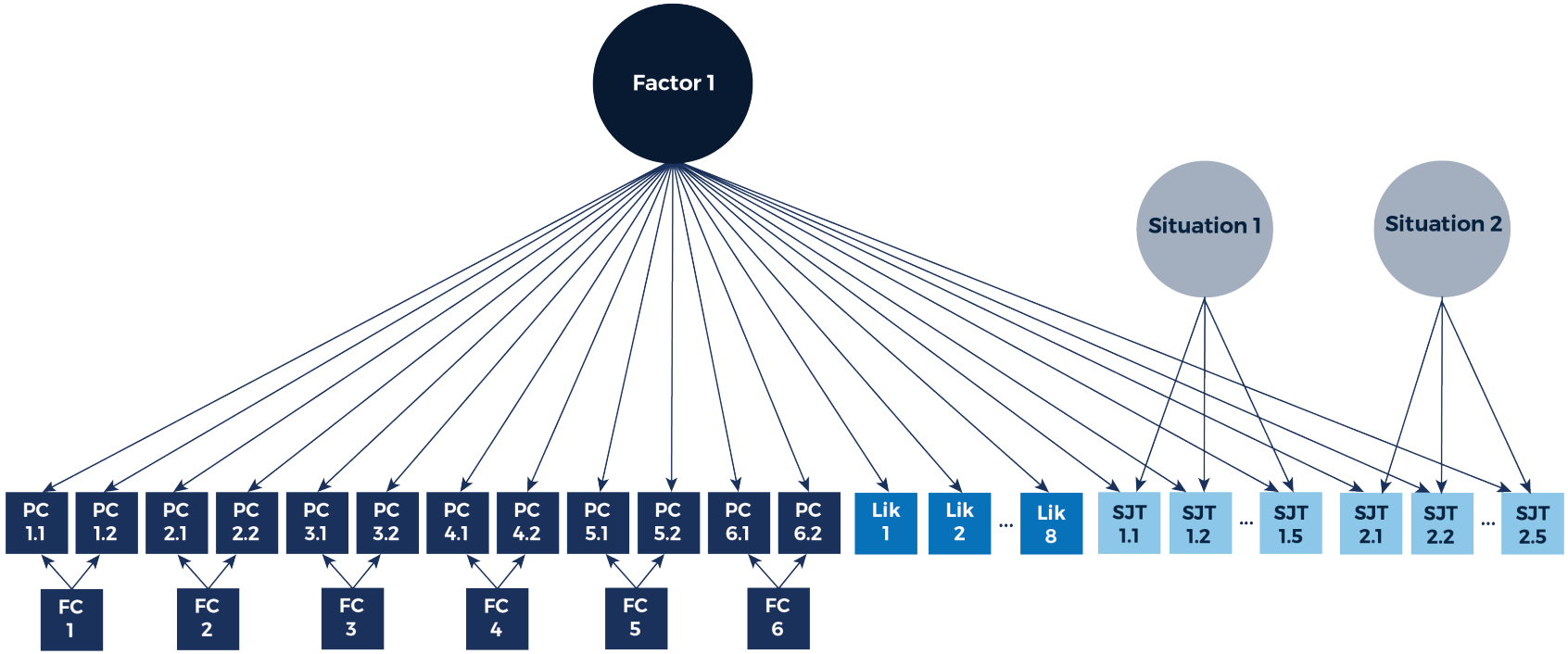
Collaborative for Academic, Social, and Emotional Learning (CASEL) Core Competency Composite Scores

In addition to a single score per social emotional skill, school aggregate and student roster reports include five composite scores that align with Collaborative for Academic, Social, and Emotional Learning (CASEL) core competencies – self-awareness, self-management, social awareness, relationship skills, and responsible decision making (CASEL, 2020). Five subject matters experts agreed upon two skills that align with each CASEL competency (see Table 3; for full methodology, see Walton, Burrus, Anguiano-Carrasco, Way, & Murano, 2019). The two skills are equally weighted by taking the mean of the two unified scores. For ease of interpretation, this mean is then transformed to a T score. Means and standard deviation to compute the T scores were obtained from the samples reported on below in sections 5.2 and 5.10.

Table 3. Alignment of Assessment Skills to the CASEL Core Competencies

CASEL Competency	Assessment Skill
Self-awareness	• Maintaining Composure • Social Connection
Self-management	• Sustaining Effort • Maintaining Composure
Social awareness	• Getting Along with Others • Keeping an Open Mind
Relationship skills	• Getting Along with Others • Social Connection
Responsible decision making	• Sustaining Effort • Getting Along with Others

Figure 5. Unified Scoring Model



Note: FC 1 = an item in a forced choice triad that is compared with the other two items in that triad, yielding two binary pairwise comparisons for that item (PC 1.1 and PC 1.2). Lik 1 = a Likert item. SJT 1.1 = a behavioral response to a single situational judgment test.

Chapter 4

Initial Item Development

Initial Item Development

To generate the initial item pool for Mosaic by ACT: Social Emotional Learning Assessment, professional item writers were provided assessment construct definitions written by research scientists (all of whom had PhDs in psychology) working on the development of the assessment. Items were contextualized so that scenarios portrayed would fit either a middle school or high school student's experience. Research scientists reviewed each item, and items were then revised by the item writers based on the feedback provided. Additional item revisions or selections were made based on data collected from 1,654 middle school and 2,105 high school students who completed an earlier version of the assessment. Items were removed if they exhibited poor psychometric qualities (e.g., they failed to load highly on their target factor). The Relationships and Safety school climate scales were taken from ACT's previous social emotional skills measure, ACT® Engage® (for information on the development of those scales, see ACT, 2016). Final Flesch-Kincaid reading levels are 5.2 and 5.8 for the Middle School assessment and the High School assessment, respectively.

Item Parameter Estimation

To estimate item parameters (i.e., discrimination and difficulty levels) for scoring purposes, large samples of middle and high school students were collected. After examining data quality and removing participants who demonstrated inattentive or biased response patterns (i.e., completing the assessment too quickly, or extremely low or high variance in their responses), samples included 11,867 middle school and 4,976 high school students. The confirmatory factor model (see Unified Score in Chapter 3) was fit to the middle and high school forms to estimate the parameters needed for the final MIRT model. As discussed in Chapter 3, these parameters were transformed into MIRT parameters. That is, for Likert and SJT items, loadings were transformed into discrimination parameters by multiplying each path coefficient by 1.702, and thresholds were transformed into difficulty parameters by multiplying by -1.702. In the case

of forced choice items, the parameters were estimated as discussed by Brown and Maydeu-Olivares (2011). Most scores obtained through the MIRT model range between -1.00 and +1.00.

Final Item Set

The final item sets for the middle and high school forms includes 40 Likert items (eight per construct), 10 SJTs (two per construct), and 10 FC triads (30 items total with six per construct). The FC triads are balanced to ensure all possible triads are represented. There are 12 Relationship climate items and 11 Safety climate items, all of which are Likert items.

Chapter 5

Reliability and Validity Evidence

Procedure

Below we report reliability and validity evidence collected from students at schools who administered Mosaic by ACT: Social Emotional Learning Assessment during the 2018-2019 academic year. All schools reported on below completed the assessment by June 2019. Details of the student characteristics and psychometric properties of the scales also appear. Two forms were administered – Middle School, which is designed for students in grades 6-8, and High School, which is designed for students in grades 9-12. Details of each are reported separately.

Students took the assessment online during a class session that was designated by their schools. Schools allotted one class period for completion of the assessment, but additional time was granted if necessary. For middle school students, the median number of minutes required to complete the assessment was 27.8, and the mode was 21.6. For high school students, the median time for completion was 23.4 minutes, and the mode was 21.7 minutes. Prior to test administration, schools scheduled test administration times, and test administrators provided schools with student login information. Students received immediate feedback, and schools received a student roster report immediately. Later in the school year, schools received an aggregate school report. Schools reported their students' grade level and gender.

A subset of four school districts reported additional middle school student data including number of absences, number of disciplinary infractions, race/ethnicity, and free/reduced lunch status. A subset of 12 districts reported these data (as well as student GPA) for high school students. One of these districts also reported student scores on the ACT® Aspire® Assessment System (ACT, 2019a), which measures student academic performance in English, reading, math, science, and writing. The primary purpose of collecting these additional data was to further evaluate the test-criterion validity of the assessment scales and to further examine subgroup differences. Some data were missing for some schools or students so not all Ns are identical across analyses.

Middle School Assessment

Method

Participants

A total of 24,400 students from 160 schools participated. The grade level breakdown is as follows: 6th grade = 3,864 (15.8%), 7th grade = 17,585 (72.1%), and 8th grade = 2,951 (12.1%). The gender breakdown is as follows: female = 12,273 (50.3%) and male = 12,127 (49.7%).

Data Exclusions

Students with inconsistent (high variance) and/or non-varied (low variance) response patterns were flagged. High variance likely indicates careless responding, while low variance likely indicates acquiescent responding or some other response bias. There were 15 possible response patterns that could be flagged across item types and assessment skills (e.g., low variance across all Likert items, high variance across Sustaining Effort SJTs), and if a student had two or more flags, he or she was not scored. In a previous study used to establish exclusion rules (N = 14,388), this led to 2.5% of the sample being excluded.

Results

Evidence for Reliability

The reliability (or precision) of an assessment refers to the consistency of scores obtained from the assessment (Standards for Educational and Psychological Testing, below referred to as the Standards; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Reliability evidence is always an important factor to consider in test development, and the need for precision increases when high-stakes decisions are made from test score interpretation. Three types of reliability evidence were collected: internal consistency reliability, empirical marginal reliability, and test-retest reliability. The three lines of reliability evidence are discussed in turn below.

Internal Consistency Reliability

Internal consistency was measured with Cronbach's alpha. Cronbach's alpha values for the two climate scales (the only scores based solely on Likert items) were as follows: Relationships = .86 and Safety = .84.

Empirical Marginal Reliability

Empirical marginal reliability estimates are a measure of the precision of the unified scores. Marginal reliability shows the proportion of variance in the observed score due to true score (equal to observed minus error variance; Green, Bock, Humphreys, Linn, & Reckase, 1984). When the variance is computed by averaging the squared standard errors of estimated trait scores in a sample (in contrast to the theoretical distribution for all trait values; du Toit, 2003),it is referred to as empirical marginal reliability. These values reached: Sustaining Effort = .88, Getting Along with Others = .86, Maintaining Composure = .89, Keeping an Open Mind = .85, and Social Connection

Test-Retest Reliability

Three-hundred sixty-five of the students completed the assessment twice during the course of the school year. The mean number of days between the two administrations was 196, and the median was 202 days (approximately 6-7 months). Test-retest correlations for the five social emotional skills were as follows: Sustaining Effort = .63, Getting Along with Others = .62, Maintaining Composure = .66, Keeping an Open Mind = .55, and Social Connection = .65. For the two climate dimensions, test-retest correlations were: Relationships = .74 and Safety = .68.

Evidence for Validity

Validity is defined as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA, APA, & NCME, p. 11). Validity can be thought of as a unitary concept with multiple sources of evidence that can support the intended interpretation of test scores for their proposed use. Two types of validity evidence were reviewed, as outlined by the Standards. These include evidence based on content and test-criterion validity. The two types of evidence are discussed in turn below.

Content Validity Evidence

Standard 1.11 of the Standards discusses the need to establish content-oriented validity evidence (AERA, APA, & NCME, 2014). Evidence based on test content validates the relationship between the test’s actual content and the constructs that the test intends to measure. This can include both logical and empirical analysis of how well the test items map to the constructs they are intended to measure. Likert, FC, and SJT items were mapped onto the social emotional skill content domain by five independent raters. The raters, all of whom were psychology doctoral students or held a PhD in psychology, reviewed the construct definitions and selected which of the five constructs they believed each item represented. The intraclass correlation coefficients (two-way random with absolute agreement) for the Likert, FC, and SJT items reached .87, .97, and .57, respectively. The intraclass correlation coefficient for the climate items reached .88. For the Likert items, four or five of the raters correctly mapped 32 of the 40 items. For the FC items, four or five of the raters correctly mapped all 30 items. Four or five of the raters correctly mapped nine of the ten SJTs. Twenty-two of the 23 climate items were correctly mapped by four or five raters.

Test-Criterion Validity Evidence

An additional standard outlined in the Standards (AERA, APA, & NCME, 2014) concerns the examination of evidence regarding relationships with criteria. These criteria performance measures are separate constructs that are hypothesized to be related to constructs measured by the test. Correlations between the five social emotional skills and the two climate scales are reported in Table 4. As expected (Allen et al., 2019; Osher & Berg, 2017), they were positively correlated.

Table 4. Correlations Between Assessment Skills and School Climate Dimensions

	Relationships	Safety
Sustaining Effort	.50*	.38*
Getting Along with Others	.50*	.38*
Maintaing Composure	.54*	.43*
Keeping an Open Mind	.48*	.34*
Social Connection	.49*	.29*

Note. N = 24,400. *p < .05.

Associations with GPA

In addition, correlations between a 12-category self-reported GPA variable (e.g., A+, 97-100%; A, 93-96%; etc.) and the five social emotional skills and two climate dimensions were examined and are reported below in Table 4. Note that Poropat (2019) reports meta-analytic estimates for primary and secondary education. Given that middle school falls between these, Poropat's primary and secondary estimates are both provided in Table 4. In line with his findings, Sustaining Effort had a strong relationship with GPA. In some cases, the assessment scales outperformed expectations based on prior literature. A regression model was fit with self-reported GPA as the outcome variable and the five scale scores as the predictor variables. The social emotional scales accounted for a significant amount of variance in GPA, $R^2 = .15$, $F(5, 22{,}777) = 817.71$, $p < .01$.

The positive correlations between GPA and Relationships and Safety (Table 5) replicate prior research suggesting a positive association between academic performance and school climate (Berkowitz, Moore, Astor, & Benbenishty, 2017). Although the F test for change in R^2 was statistically significant (likely due to the large sample size and high power), no additional variance (i.e., less than 1%) was accounted for when entering climate in a second step in a hierarchical regression model, $\Delta R^2 = .00$, $F(2, 22{,}775) = 5.16$, $p = .01$. That is, school climate explained no incremental variance in GPA beyond that accounted for by the social emotional skills.

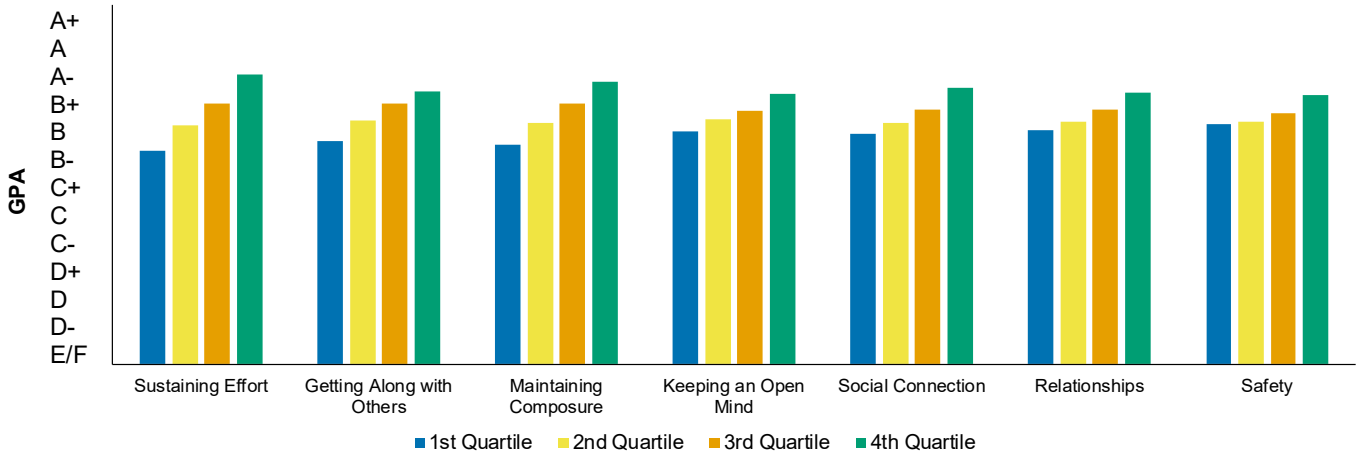
Table 5. Correlations Between Middle School Social Emotional Skills and School Climate Dimensions and GPA

	Poropat's (2009) Meta-Analysis		With Self-Reported GPA[a]
	Primary Education	Secondary Education	
Sustaining Effort (Conscientiousness)	.28	.21	.38*
Getting Along with Others (Agreeableness)	.30	.05	.23*
Maintaining Composure (Emotional Stability)	.20	.01	.30*
Keeping an Open Mind (Openness to Experience)	.24	.12	.20*
Social Connection (Extraversion)	.18	-.03	.23*
Relationships			.20*
Safety			.14*

Note. [a]*N* = 22,783. *$p < .05$.

See Figure 6 for an illustration of the relationship between social emotional skills and school climate and GPA. Individuals were identified as scoring in the 1st (bottom 25%), 2nd (second 25%), 3rd (third 25%), or 4th (top 25%) quartile on each construct. Average self-reported GPA, which ranged from 1 (E/F, below 65%) to 12 (A+, 97-100%), was calculated for students within each quartile. For example, individuals falling in the bottom quartile on Sustaining Effort have an average GPA of slightly greater than a B- (80-82%), while individuals falling in the top quartile on these skills have an average GPA of nearly an A- (90-92%). That is, individuals with higher Sustaining Effort score roughly a full letter grade higher than individuals with lower Sustaining Effort .

Figure 6. Average Self-Reported GPA by Quartiles of Middle School Social Emotional Skills and School Climate Scores



Associations with Absences

School-reported absences were available for 294 students. There are inconsistent findings on the associations between student characteristics and absenteeism reported in the literature. For example, Lounsbury, Steel, Loveland, and Gibson (2004) reported that, of the Big Five, openness had the strongest association with absenteeism among 7th and 12th graders and the second strongest association with absenteeism among 10th graders. For these primary and secondary students, openness had a negative association with absenteeism. However, the opposite effect has been reported among college students. Chamorro-Premuzic and Furnham (2003) also found the strongest relationship between absenteeism and openness among college students, but they were positively correlated with one another. Credé, Roch, and Kieszczynka (2010) carried out a meta-analysis including a look at student characteristic-attendance relationships among college students and, in general, reported weak to moderate associations. Returning to the example of openness, they reported a near-zero association and instead reported the strongest (positive) effect for conscientiousness. We examined correlations between total number of absences (M = 8.34; SD = 7.68; minimum = 0; maximum = 48), which was a sum of excused and unexcused absences, and the five social emotional skills. The correlations were all near zero and none were significant: Sustaining Effort (r = -.01), Getting Along with Others (r = -.00), Maintaining Composure (r = .06), Keeping an Open Mind (r = .01), and Social Connection (r = .01). Climate was also unrelated to absences: Relationships (r = .02) and Safety (r = -.02). In terms of school climate, prior research shows that schools with poorer school climate generally have greater problems of absenteeism (ACT, 2016; Van Eck, Johnson, Bettencourt, & Johnson, 2017), though we found no association between climate and absenteeism. The social emotional skills did not account for a significant amount of variance in absences, R^2 = .01, $F(5, 288)$ = .53, p = .76, nor did climate when entered into the hierarchical regression model, ΔR^2 = .00, $F(2, 286)$ = .26, p = .77.

Associations with Discipline

Table 6 provides the correlations between the assessment scales and number of reported disciplinary infractions (M = 1.30; SD = 3.13; minimum = 0; maximum = 25). Prior research on behavior problems in children and adolescents suggests Sustaining Effort, Getting Along with Others, and Maintaining Composure should have negative associations with discipline, and Keeping an Open Mind and Social Connection should have positive, yet smaller associations (Tackett, 2006; Tackett, Kushner, De Fruyt, & Mervielde, 2013). Correlations were largely in the expected direction, but Sustaining Effort had the strongest relationship with discipline (r = -.15) rather than Getting Along with Others like some prior research suggests (Tackett et al., 2013). A regression model was fit with number of reported disciplinary infractions as the outcome variable and the five social emotional skills as the predictor variables. The assessment scales accounted for a significant amount of variance in discipline, R^2 = .04, $F(5, 344)$ = 2.88, $p < .05$.

Perceptions of positive school climate, including better relationships at school and feelings of safety, ought to be associated with fewer behavioral problems (Osher & Berg, 2017). Although in the expected direction, the correlations between number of disciplinary infractions and Relationships (r = -.08) and Safety (r = -.03) were small. No additional variance was explained when climate was entered into a hierarchical regression model predicting number of disciplinary infractions, ΔR^2 = .00, $F(2, 342)$ = .75, p = .47.

Rather than reporting the number of disciplinary infractions, some schools only provided a binary response (i.e., no infractions vs. at least one). We dichotomized the continuous responses some schools provided and combined them with the binary responses and carried out an independent samples t-test to compare these students. These results and standardized effect sizes are reported in Table 5. Students with no disciplinary infractions scored higher on all social emotional skills with the strongest effect for Sustaining Effort, d = .55. In general, there was a negative relationship between discipline rates and school climate, though the only test that was statistically significant was the test comparing Relationships between those with no disciplinary infractions and those with at least one, d = .33.

Table 6. Associations Between Middle School Social emotional Skills and School Climate Dimensions and Discipline

		0 Discipline Problems[b]		≥ 1 Discipline Problem[c]			
	r[a]	*M*	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	-.15*	.13	1.14	-.46	.93	5.81*	.55
Getting Along with Others	-.06	.34	1.46	-.13	1.90	3.13*	.30
Maintaining Composure	-.11*	.10	1.18	-.30	1.29	3.50*	.33
Keeping an Open Mind	-.02	.06	1.40	-.29	1.22	2.76*	.26
Social Connection	.03	-.00	1.24	-.22	1.18	1.95	.18
Relationships	-.08	4.12	.86	3.85	.84	3.45*	.33
Safety	-.03	4.27	.80	4.15	.83	1.63	.15

Note. [a]N = 350. [b]N = 335. [c]N = 169. *p < .05.

Subgroup Differences

Finally, demographic group differences were evaluated to determine whether there were any significant subgroup differences on the assessment scales. Independent samples *t*-tests were carried out to compare students who were identified as male vs. female (self-reported), free/reduced lunch status versus not (school-reported), and those who identified as White versus an underrepresented minority (URM) group (school-reported). The URM group included students who identified as any racial or ethnic group other than White or Asian. This included American Indian, Black, Hispanic, and bi/multiracial students. Asian students typically are not underrepresented in higher education settings, and there were only three Asian students in this dataset, so they were not included in either of the two groups. Descriptive statistics for the scales, as well as results of the *t*-tests and standardized effect sizes, can be found in Tables 7-9.

Gender Differences

De Bolle et al. (2015) reported findings from a large cross-cultural study of gender differences in adolescence. From ages 12-14, females scored significantly higher than males on each of the Big Five (see above for Big Five-assessment alignment) with the exception of emotional stability. The most consistent and robust findings were for conscientiousness and openness, with all standardized effect sizes exceeding .25 for openness and .30 for conscientiousness. At age 12, females scored higher than males on emotional stability but to a trivial degree, *d* = .01. By age 13, females scored lower than males on emotional stability, but the difference was not statistically significant. Our findings are largely in line with De Bolle and colleagues' in terms of direction and magnitude of observed effects. Female and male students had significantly different scores on all social emotional skills with females scoring higher on all constructs (see Table 7). Effect sizes ranged from .07 (Maintaining Composure) to .31 (Getting Along with Others). Gender differences on climate were statistically significant yet slight with effect sizes less than +.04. Previous research shows that males typically report lower climate scores than females (Buckley, Storino, & Sebastiani, 2003; Koth, Bradshaw, & Leaf, 2008).

Table 7. Gender Differences on Middle School Assessment Scales

	Female[a]		Male[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	.12	1.28	-.18	1.17	19.06*	.24
Getting Along with Others	.47	1.74	-.06	1.62	24.57*	.31
Maintaining Composure	.04	1.38	-.05	1.33	5.17*	.07
Keeping an Open Mind	.18	1.26	-.14	1.29	19.56*	.25
Social Connection	.07	1.32	-.12	1.34	10.92*	.14
Relationships	3.98	.87	3.95	.84	2.69*	.03
Safety	4.25	.89	4.28	.87	-2.70*	-.03

Note. *$p < .05$. [a]N = 12,273. [b]N = 12,127.

Free/Reduced Lunch Status Differences

Free/reduced lunch status was the only index available related to socioeconomic status (SES). Meta-analytic data suggest that parental SES has weak, positive associations with young adults' Big Five personality traits, with the strongest observed effect size for openness at .14 (Ayoub, Gosling, Potter, Shanahan, & Roberts, 2018). However, we observed the opposite direction in our data set; students who are eligible for free/reduced lunches scored significantly higher on Keeping an Open Mind, *d* = -.19. There were no other significant social emotional skill differences observed and no other sizable effects

(see Table 8). Students eligible for free/reduced lunch also scored significantly higher on Safety, *d* = -.23. This differs from some previous research showing that higher concentrations of school poverty in student populations are negatively related to climate scores (Allen et al., 2019). However, students who are provided food by the school probably feel safer and more supported at school, especially if they would not have had food otherwise. The results of a recent meta-analysis on climate and SES showed that a positive school climate can lessen the negative effects of low SES on academic achievement. The moderating influence of school climate was particularly strong for students from low SES backgrounds, such that these students do better academically when in a positive school climate (Berkowitz, Moore, Astor, & Benbenishty, 2017).

Table 8. Free/Reduced Lunch Status Differences on Middle School Assessment Scales

	Ineligible for Free/Reduced Lunch[a]		Eligible for Free/Reduced Lunch[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	.03	1.19	-.12	-.12	1.37	.13
Getting Along with Others	.22	2.05	.17	.17	.37	.03
Maintaining Composure	-.04	1.26	-.03	-.03	-.11	-.01
Keeping an Open Mind	-.24	1.32	.02	.02	-1.99*	-.19
Social Connection	-.05	1.28	-.09	-.09	.28	.03
Relationships	4.10	.92	4.00	.83	1.18	.11
Safety	4.09	.85	4.29	.79	-2.47*	-.23

Note. *$p < .05$. [a]$N = 153$. [b]$N = 351$.

Race/Ethnicity Differences

According to Foldes, Duehr, and Ones's (2008) meta-analytic data, White and Black students generally show negligible differences 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). There were fairly small effect sizes observed in our data, which is consistent with Foldes and colleagues' findings. URM students scored higher than White students on all social emotional skills and climate dimensions, though the only difference that was statistically significant was Keeping an Open Mind (see Table 9). Effect sizes ranged from -.03 (Relationships) to -.18 (Getting Along with Others and Keeping an Open Mind). There were no significant differences on the climate dimensions. This differs from previous research in that White students typically report higher climate scores as compared to minorities (Koth et al., 2008; Mitchell, Bradshaw, & Leaf, 2010).

Table 9. Race/Ethnicity Differences on Middle School Assessment Scales

	White[a]		Underrepresented Minority[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	-.12	1.16	-.02	1.03	-.98	-.09
Getting Along with Others	.07	1.36	.36	1.97	-1.97	-.18
Maintaining Composure	-.12	1.18	.09	1.32	-1.87	-.17
Keeping an Open Mind	-.16	1.44	.08	1.21	-1.97*	-.18
Social Connection	-.12	1.27	.01	1.15	-1.16	-.11
Relationships	4.02	.91	4.05	.78	-.38	-.03
Safety	4.18	.83	4.30	.78	-1.64	-.15

Note. *$p < .05$. [a]$N = 299$. [b]$N = 200$.

Norms

In reports provided to students, scores are conveyed using a four-point rating system that mimics a gas gauge. The lowest level of the gauge, which is described as “developing,” corresponds to a raw score of -1 or lower, which is one standard deviation below 0 (so this corresponds to the approximately the 16th percentile or below). The second, “approaching,” corresponds to a raw score between -.9999 and 0 (approximately the 17th-50th percentiles). The third, “demonstrating,” corresponds to a raw score between .0001 and 1 (approximately the 51st-84th percentiles). The highest level of the gauge, “mastering,” corresponds to a raw score greater than 1 (approximately at or above the 85th percentile). These ratings are based on data from the pilot study sample (i.e., the sample of 11,867 students mentioned in Chapter 4).

High School Assessment

Most procedures and analyses were identical to those discussed above for Middle School; therefore, we primarily report only statistics in this section. Any deviations or additional analyses are explained below in detail.

Method

Participants

A total of 9,112 students from 93 schools participated. The grade level breakdown is as follows: 9th grade = 5,413 (59.4%), 10th grade = 1,739 (19.1%), 11th grade = 1,002 (11.0%), and 12th grade = 958 (10.5%). The gender breakdown is as follows: female = 4,792 (52.6%) and male = 4,320 (47.4%).

Data Exclusions

Flagging procedures were the same as that for Middle School. In a previous sample used to establish exclusion rules (*N* = 5,575), using this method led to 1.4% of the sample being excluded.

Results

Evidence for Reliability

Internal consistency reliability

Cronbach's alpha reached .88 for Relationships and .85 for Safety.

Empirical marginal reliability

Estimated marginal reliability estimates reached: Sustaining Effort = .90, Getting Along with Others = .89, Maintaining Composure = .86, Keeping an Open Mind = .85, and Social Co*nnection* = .88.

Test-retest reliability

One hundred eighty-seven of the students reported on above completed the assessment twice during the course of the school year. The mean number of days between Time 1 and Time 2 was 142, and the median was 152 days (approximately 5 months). Test-retest correlations were as follows: Sustaining Effort = .76, Getting Along with Others = .64, Maintaining Composure = .58, Keeping an Open Mind = .68, and Social Connection = .65. For the two climate dimensions, test-retest correlations were: Relationships = .68 and Safety = .72.

Evidence for Validity

Content validity evidence

As with Middle School, a panel of subject matter experts reviewed the construct definitions and selected which of the five constructs they believed each item represented. The intraclass correlation coefficient for the Likert items and SJTs reached .77 and .86, respectively. Four or five of the raters correctly mapped 31 of 40 Likert items. Nine of the ten SJTs were correctly mapped by four or five raters. The FC and climate items on Middle School and High School are identical.

Test-criterion validity evidence

Correlations between the five social emotional skills and the two climate scales are presented in Table 10. As expected, all are positive (Allen et al., 2019; Osher & Berg, 2017).

Table 10. Correlations Between High School Social Emotional Skills and School Climate Dimensions

	Relationships	Safety
Sustaining Effort	.41*	.29*
Getting Along with Others	.44*	.33*
Maintaining Composure	.41*	.32*
Keeping an Open Mind	.39*	.24*
Social Connection	.42*	.19*

Note. N = 9,112. *p < .05.

Associations with GPA

Correlations between a 12-category self-reported GPA variable (e.g., A+, 97-100%; A, 93-96%; etc.) and the five social emotional skills and climate were examined and are reported in Table 10. All were positive and significant and, in line with Poropat's (2009) meta-analytic findings, Sustaining Effort had the strongest relationship with GPA. In all cases, the scales outperformed expectations based on prior literature (Poropat, 2009). A regression model was fit with self-reported GPA as the outcome variable and the five social emotional skills as the predictor variables. The scales accounted for a significant amount of variance in GPA, R^2 = .24, F(5, 8,876) = 553.30, p < .01. Although the F test for change in R^2 was statistically significant (likely due to the large sample size and high power), no additional variance was accounted for when entering climate in a second step in a hierarchical regression model, ΔR^2 = .00, F(2, 8,874) = 15.09, p < .01.

Table 11. Correlations Between High School Social Emotional Skills and School Climate Dimensions and GPA

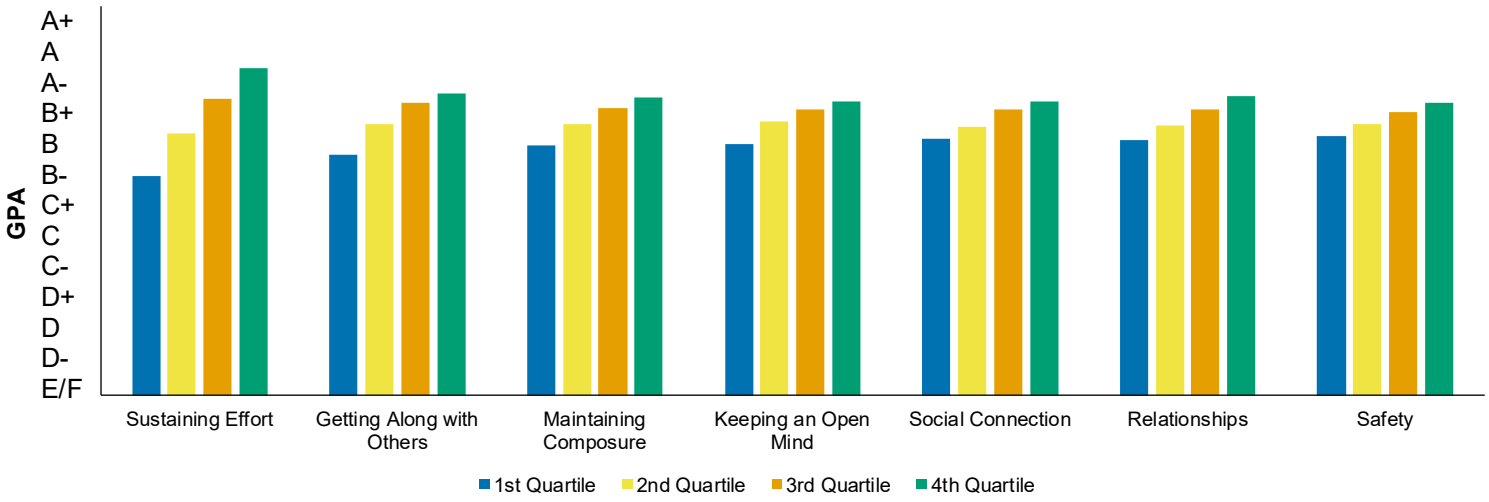
	Poropat's (2009) Meta-Analysis	With Self-Reported GPA[a]	With School-Reported GPA[b]
Sustaining Effort (Conscientiousness)	.21	.48*	.41*
Getting Along with Others (Agreeableness)	.05	.27*	.29*
Maintaining Composure (Emotional Stability)	.01	.21*	.18*
Keeping an Open Mind (Openness to Experience)	.12	.20*	.17*
Social Connection (Extraversion)	-.03	.18*	.11*
Relationships		.21*	.13*
Safety		.17*	.14*

Note. [a]N = 8,882. [b]N = 1,058. *p < .05.

Table 11 also provides the correlations between the assessment scales and school-reported GPA in the subsample. For the 1,035 students with both self- and school-reported GPA, the correlation between the two reached .76. As expected (Poropat, 2009), Sustaining Effort has the strongest association with GPA. A regression model was fit with school-reported GPA as the outcome variable and the five social emotional skills as the predictor variables. The five skills accounted for a significant amount of variance in GPA, $R^2 = .19$, $F(5, 1{,}052) = 50.06$, $p < .01$. Adding school climate in a second step of a hierarchical regression did not result in additional variance accounted for, $\Delta R^2 = .00$, $F(2, 1{,}050) = .22$, $p = .15$.

See Figure 7 for a graphical depiction of the self-report findings. Individuals were identified as scoring in the 1st (bottom 25%), 2nd (second 25%), 3rd (third 25%), or 4th (top 25%) quartile on each construct. These quartiles were compared in terms of their self-reported GPA, which ranged from 1 (E/F, below 65%) to 12 (A+, 97-100%). For example, individuals falling in the bottom quartile on Sustaining Effort have an average GPA of less than a B- (80-82%), while individuals falling in the top quartile on these skills have an average GPA of over an A- (90-92%). That is, individuals high on Sustaining Effort score more than a full letter grade higher than individuals low on Sustaining Effort.

Figure 7. Average Self-Reported GPA by Quartiles of High School Social Emotional Skills and School Climate Scores



Associations with Absences

School-reported absences were available for 890 students. We examined correlations between total number of absences ($M = 8.64$; $SD = 11.18$; minimum = 0; maximum = 101.5), which was a sum of excused and unexcused absences, and the five social emotional skills and two climate dimensions (see second column of Table 12). A greater number of absences was significantly associated with lower Sustaining Effort, Getting Along with Others, and Maintaining Composure. The five social emotional skills together accounted for a significant amount of variance in absenteeism, $R^2 = .03$, $F(5, 884) = 5.67$, $p < .01$. No additional variance in absenteeism was explained by adding the climate dimensions to the model, $\Delta R^2 = .00$, $F(2, 882) = .78$, $p = .46$. The correlations were of greater magnitude than in the middle school sample, which likely reflects greater autonomy among high school students relative to middle school students. That is, student characteristics likely play a greater role in behaviors such as truancy,

whereas in middle school, student absenteeism is more likely to reflect family characteristics. Although inconsistent findings are reported in the literature and it is therefore difficult to evaluate our findings in relation to those in the extant literature, what is important to note is that High School social emotional skills explain a significant amount of variance in absenteeism.

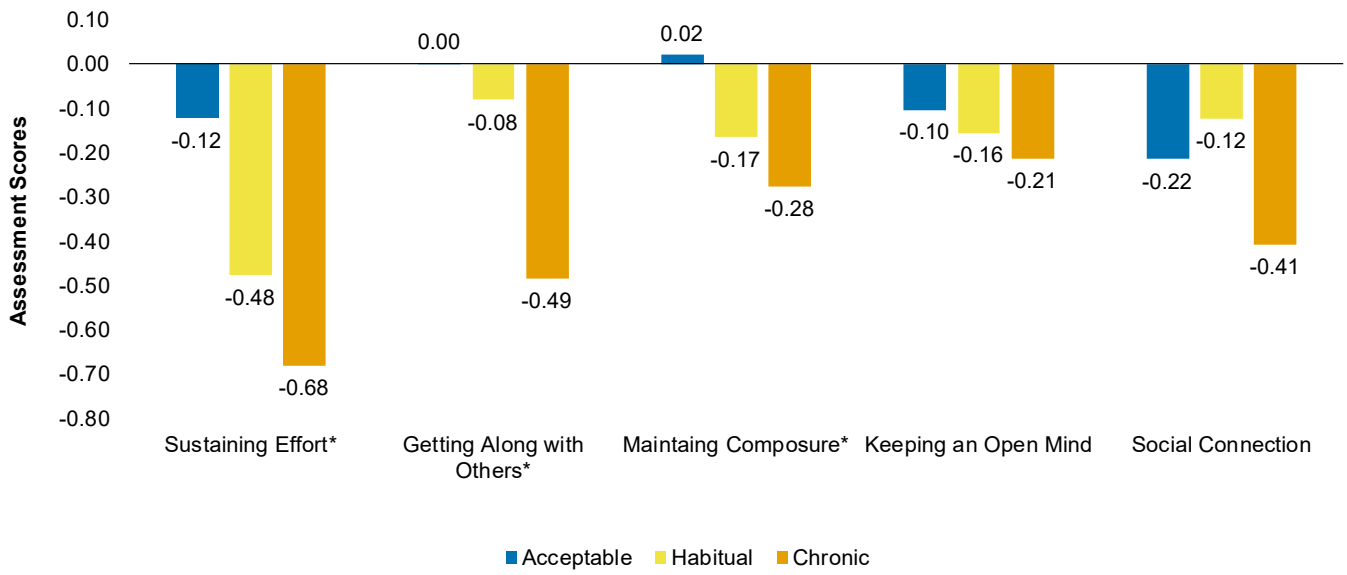
Table 12. Associations between High School Scales and Climate and Absenteeism

		Acceptable[b]		Habitual[c]		Chronic[d]		
	r[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*
Sustaining Effort	-.14*	-.12	1.32	-.48	1.26	-.68	1.41	10.07*
Getting Along with Others	-.11*	.00	1.27	-.08	1.27	-.49	1.16	5.67*
Maintaining Composure	-.08*	.02	1.23	-.17	1.05	-.28	1.22	3.43*
Keeping an Open Mind	-.03	-.10	1.30	-.16	1.23	-.21	1.38	.34
Social Connection	-.02	-.22	1.27	-.13	1.32	-.41	1.24	.24
Relationships	-.02	3.82	.88	3.64	.92	3.74	.90	2.97
Safety	-.02	4.22	.79	4.07	.85	4.08	.94	3.04*

Note. [a]*N* = 890. [b]*N* = 636. [c]*N* = 168. [d]*N* = 86. *F* df = 2, 887. *$p < .05$.

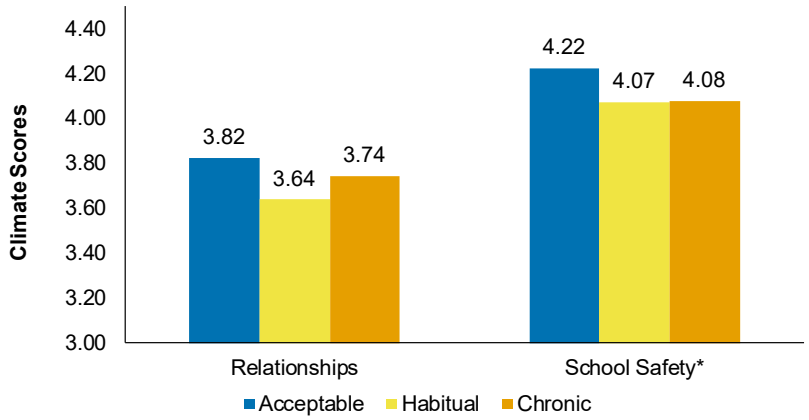
We also split the sample into three groups representing students with acceptable absentee records (fewer than 10 missed days), habitual absentee records (defined as at least 10 missed days but fewer than 18), or chronic absentee records (defined as 18 or more missed days) and examined group mean differences. Most states describe chronic absenteeism as missing 10% or more days within a school year (Attendance Works, n.d.), which would equate to 18 or more days, and some states consider missing 10 or more days within a school year as being habitually truant (Colorado Department of Education, 2018). The chronic group had the lowest means on all skills, and there were statistically significant group differences on Sustaining Effort, Getting Along with Others, and Maintaining Composure, as well as Safety. Post-hoc analyses showed that the acceptable group scored significantly higher than the habitual and chronic groups on Sustaining Effort, the chronic group scored significantly lower than the acceptable and habitual groups on Getting Along with Others, the acceptable group scored significantly higher than the chronic group on Maintaining Composure, and the acceptable group scored significantly higher than the habitual group on Safety. See Figures 8 and 9 for a depiction of the differences on the social emotional skills and climate.

Figure 8. Associations between High School Social Emotional Skills and Absenteeism



Note: *p <.05.

Figure 9. Associations between High School Climate Dimensions and Absenteeism



Note: *p <.05.

Associations with Discipline

Table 13 provides the correlations between the assessment scales and number of reported disciplinary infractions (M = 3.26; SD = 6.13; minimum = 0; maximum = 54). Social Connection had the strongest relationship with discipline (r = .12). A regression model was fit with number of reported disciplinary infractions as the outcome variable and the five social emotional skills as the predictor variables. The skills accounted for a significant amount of variance in discipline, R^2 = .07, $F(5, 715) = 9.88$, $p < .01$.

Entering the two climate dimensions into the second step of a hierarchical regression led to a significant increase in variance explained, $\Delta R^2 = .03$, $F(2, 713) = 12.40$, $p < .01$.

Results of the dichotomized (i.e., no infractions vs. at least one) discipline variable are reported in Table 13. There were significant mean group differences on Sustaining Effort ($d = .17$), Getting Along with Others ($d = .23$), Maintaining Composure ($d = .13$), Social Connection ($d = -.22$), Relationships ($d = .13$), and Safety ($d = .18$). With the exception of Social Connection, the group with no discipline problems exhibited better social emotional skills and better perceptions of school climate. Any differences in direction of the effect across binary and continuous analyses, which occurred specifically for Keeping an Open Mind and Safety, are due to non-identical samples used across the analyses.

Table 13. Associations Between High School Social Emotional Skills and School Climate Dimensions and Discipline

		0 Discipline Problems[b]		≥1 Discipline Problem[c]			
	r[a]	*M*	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	-.05	.02	1.29	-.19	1.28	3.11*	.17
Getting Along with Others	-.09*	.21	1.21	-.07	1.28	4.16*	.23
Maintaining Composure	-.05	.13	1.20	-.02	1.22	2.30*	.13
Keeping an Open Mind	.05	.14	1.25	.07	1.33	.99	.05
Social Connection	.12*	-.23	1.26	.05	1.26	-4.05*	-.22
Relationships	-.08*	3.89	.80	3.78	.90	2.34*	.13
Safety	.08*	4.35	.73	4.21	.80	3.26*	.18

Note. [a]$N = 721$. [b]$N = 982$. [c]$N = 508$. *$p < .05$.

Subgroup Differences

Finally, demographic group differences were evaluated to determine whether there were any significant subgroup differences on the assessment scales. In addition to the independent samples *t*-tests described above, subgroup differences on the ACT® Aspire® Assessment System were examined for one school who provided those data on their 9th-grade students to compare the magnitude of differences on academic content versus social emotional skills. We report on the ACT Aspire Interim Composite score (i.e., the mean score of the four subject area tests) only as all subtests yielded the same pattern of results.

Gender Differences

De Bolle et al. (2015) reported findings from a large cross-cultural study of gender differences in adolescence. From ages 15-17, females scored significantly higher on conscientiousness, emotional stability, and openness to experience (see above for crosswalk between Big Five and assessment constructs). Females scored significantly higher than males on extraversion at ages 15 and 16 and significantly higher on agreeableness at age 17. Emotional stability (Maintaining Composure) shows the most interesting developmental pattern with females scoring slightly higher than, or equal to, males in early adolescence, but then males surpassing females in middle-late adolescence, with that difference maintaining through adulthood (De Bolle et al., 2015). In terms of social emotional skills, female students scored significantly higher on Sustaining Effort, Getting Along with Others, and Keeping an Open Mind, with standardized effect sizes ranging from .24 (Keeping an Open Mind) to .46 (Getting Along with Others; see Table 14). Male students scored significantly higher on Maintaining Composure, d = -.12. These findings are largely in line with previous findings in terms of direction and magnitude of observed effects. Moreover, the developmental pattern for Maintaining Composure is largely replicated across the middle and high school forms of the assessment.

Male students scored significantly higher than female students on Safety, d = -.07. Male and female students did not score significantly different from one another on ACT Aspire.

Table 14. Gender Differences on the High School Assessment and ACT Aspire

	Female[a]		Male[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	.17	1.38	-.35	1.30	18.58*	.39
Getting Along with Others	.39	1.35	-.23	1.27	22.24*	.46
Maintaining Composure	-.03	1.22	.13	1.34	-5.66*	-.12
Keeping an Open Mind	.21	1.24	-.09	1.33	11.23*	.24
Social Connection	-.14	1.28	-.18	1.24	1.56	.03
Relationships	3.92	.85	3.91	.86	.75	.02
Safety	4.26	.84	4.31	.80	-3.19*	-.07
ACT Aspire	162.30[c]	3.48	162.61[d]	3.84	-.74	-.09

Note. *p < .05. [a]N = 4,792 for all variables except ACT Aspire. [b]N = 4,320 for all variables except ACT Aspire. [c]N = 142. [d]N = 173.

Free/Reduced Lunch Status Differences

When comparing students eligible for free/reduced lunch with ineligible students, there were no significant differences observed and no sizable effect sizes (see Table 15). The school with available ACT Aspire scores did not have available data on free or reduced lunch eligibility.

Table 15. Free/Reduced Lunch Status Differences on the High School Assessment

	Ineligible for Free/Reduced Lunch[a]		Eligible for Free/Reduced Lunch[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	-.02	1.35	-.09	1.37	.81	.05
Getting Along with Others	.10	1.26	.02	1.37	.99	.06
Maintaining Composure	.05	1.17	-.09	1.26	1.81	.11
Keeping an Open Mind	.02	1.25	.12	1.40	-1.23	-.08
Social Connection	-.11	1.25	-.02	1.30	-1.18	-.07
Relationships	3.82	.92	3.77	.90	1.01	.06
Safety	4.34	.80	4.24	.84	1.93	.12

Note. $^{*}p < .05$. [a]$N = 570$. [b]$N = 474$.

Race Differences

Underrepresented minority students scored higher than White students on all constructs, though the only difference that was statistically significant was Keeping an Open Mind, $d = -.27$ (see Table 16). URM students also scored significantly higher on Safety, $d = -.22$. White students scored higher than URM students on ACT Aspire, $d = .64$. It is typical to observe greater subgroup differences on standardized achievement tests (ACT, 2019b) than on social emotional skills assessments (Foldes et al., 2008).

Table 16. Race/Ethnicity Differences on the High School Assessment and ACT Aspire

	White[a]		Underrepresented[b]			
	M	*SD*	*M*	*SD*	*t*	*d*
Sustaining Effort	-.17	1.33	-.07	1.24	-1.63	-.08
Getting Along with Others	.02	1.32	.12	1.21	-1.57	-.08
Maintaining Composure	-.02	1.23	.07	1.18	-1.41	-.07
Keeping an Open Mind	-.09	1.27	.25	1.28	-5.38*	-.27
Social Connection	-.15	1.28	-.08	1.25	-1.08	-.05
Relationships	3.86	.90	3.80	.81	1.38	.07
Safety	4.18	.81	4.35	.72	-4.44*	-.22
ACT Aspire	163.92[c]	3.99	161.53[d]	3.15	5.88*	.64

Note. *$p < .05$. [a]$N = 922$ for all variables except ACT Aspire. [b]$N = 727$ for all variables except ACT Aspire. [c]$N = 123$. [d]$N = 190$.

Norms

The same four-point system described for Middle School students is used in the High School student reports. These ratings are based on data from the pilot study sample (i.e., the sample of 4,976 students mentioned in Chapter 4).

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*(2), 219-245.

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

ACT. (2018). *ACT® Tessera technical bulletin*. Iowa City, IA: ACT.

ACT. (2019a). *ACT® Aspire® periodic technical manual*. Iowa City, IA: ACT.

ACT. (2019b). *The ACT® 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.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i-171.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anguiano-Carrasco, C., Walton, K. E., Murano, D., Burrus, J., & Way, J. (2018). *Validity of ACT Tessera by ACT unified score*. Iowa City, IA: ACT.

Attendance Works. (n.d.). *Chronic absence*. Retrieved from https://www.attendanceworks.org/chronic-absence/the-problem/

Ayoub, M., Gosling, S. D., Potter, J., Shanahan, M., & Roberts, B. W. (2018). The relations between parental socioeconomic status, personality, and life outcomes. *Social Psychological and Personality Science*, *9*(3), 338-352.

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.

Bebell, D., & Stemler, S. (2012). *The school mission statement: Values, goals, and identities in American education*. New York, NY: Routledge.

Belfield, C., Bowden, B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). *The economic value of social and emotional learning*. New York, NY: Teachers College, Columbia University.

Berg, J., Osher, D., Moroney, D., & Yoder, N. (2017). *The intersection of school climate and social and emotional development*. Washington, DC: American Institutes for Research. Retrieved from https://www.air.org/sites/default/files/downloads/report/Intersection-School-Climate-and-Social-and-Emotional-Development-February-2017.pdf

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*(2), 425-469.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, *130*(6), 887-919.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*(3), 460-502.

Buckley, M. A., Storino, M., & Sebastiani, A. M. (2003, August). *The impact of school climate: Variation by ethnicity and gender*. Paper presented at the Annual Conference of the American Psychological Association, Toronto, ON, Canada.

Cattell, R. B. (1943). The description of personality: The foundations of trait measurement. *Psychological Review*, *50*(6), 559-594.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, *37*(4), 319-338.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*(3), 267-307.

Collaborative for Academic, Social, and Emotional Learning. (2020). *Core SEL competencies.* Retrieved from casel.org/core-competencies/

Colorado Department of Education. (2018). *Student attendance*. Retrieved from http://www.cde.state.co.us/communications/schoolattendance-factsheet-2018

Costa, P. T., Jr., & McCrae, R. R. (1994). Stability and change in personality from adolescence through adulthood. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 139-150). Hillsdale, NJ: Erlbaum.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21-50.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272-295.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, IL: University of Illinois Press.

De Bolle, M., De Fruyt, F., McCrae, R. R., Lockenhoof, C. E., Costa, P. T., Jr., Aguilar-Vafaie, M. E., … Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology*, *108*(1), 171-185.

de Raad, B. (2000). *The Big Five Personality Factors: The psycholexical approach to personality*. Kirkland, WA: Hogrefe & Huber Publishers.

de Raad, B., & Mlačić, B. (2015). The lexical foundation of the Big Five-Factor Model. In T. A. Widiger (Ed.), *The Oxford handbook of the Five Factor Model* (pp. 191-216). New York, NY: Oxford University Press.

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

Diener, E., & Lucas, R. E. (1999). Personality and subjective well-being. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 213-229). New York, NY: Russel Sage Foundation.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417-440.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions* (Tech. Rep. No. 1311). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

du Toit, M. (2003). *IRT from SSI*. Chicago, IL: Scientific Software International.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment*, *91*(2), 166-174.

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.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*, 294-309.

Dweck, C. S. (2012). *Mindset: How you can fulfil your potential*. London, England: Constable & Robinson Limited.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, *44*(3), 329-344.

Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology*, *61*(3), 579-616.

Ge, X., & Conger, R. D. (1999). Adjustment problems and emerging personality characteristics from early to late adolescence. *American Journal of Community Psychology*, *27*(3), 429-459.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26-34.

Grant, S., Hamilton, L. S., Wrabel, S. L., Gomez, C. J., Whitaker, A., Leschitz, J. T., ... & Harris, M. (2017). *Social and emotional learning interventions under the Every Student Succeeds Act: Evidence review*. *Research Report RR-2133-WF*. Santa Monica, CA: RAND Corporation.

Green, B. F., Bock, R., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347-360.

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

Hoge, D. R., Smit, E. K., & Hanson, S. L. (1990). School experiences predicting changes in self-esteem of sixth and seventh-grade students. *Journal of Educational Psychology*, *82*(1), 117-127.

Hontangas, P. M., de la Torre, J., Ponsoda, V., Leenen, I., Morillo, D., & Abad, F. J. (2015). Comparing traditional and IRT scoring of forced-choice tests. *Applied Psychological Measurement*, *39*(8), 598-612.

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 and R. Ployhart (Eds.), *Situational Judgment Tests* (pp. 205-323). Mahwah, NJ: Lawrence Erlbaum Associates.

Howard, A., & Choi, M. (2000). How do you assess a manager's decision-making abilities? The use of situational inventories. *International Journal of Selection and Assessment*, *8*(2), 85-88.

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), 839-857.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, *13*(4), 371-388.

James, W. (1981). *The principles of psychology* (Vol. 1). Cambridge, MA: Harvard University Press.

Jensen-Campbell, L. A., Adams, R., Perry, D. G., Workman, K. A., Furdella, J. Q., & Egan, S. K. (2002). Agreeableness, extraversion, and peer relations in early adolescents: Winning friends and deflecting aggression. *Journal of Research in Personality*, *36*(3), 224-251.

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, *69*(2), 323-361.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765-780.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*(3), 530-541.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*(1), 165-172.

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, *19*(1), 47-52.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology*, *100*(1), 96-104.

Kyllonen, P. C., & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. W. Engle (Eds.) *Handbook of understanding and measuring intelligence* (pp. 11-25). Thousand Oaks, CA: Sage.

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.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.). (2008). *Extending intelligence: Enhancement and new constructs*. New York, NY: Routledge.

Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for Black and White students. *Education and Treatment of Children*, *34*(2), 167-192.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*(2), 460-468.

Lipnevich, A. A., MacCann, C., & Roberts, R. D. (2013). Assessing noncognitive constructs in education: A review of traditional and innovative approaches. In D. H. Saklofske, C. R. Reynolds, & V. L.

Schwean (Eds.), *Oxford handbook of child psychological assessment*. (pp. 750-772). Cambridge, MA: Oxford University Press.

Lounsbury, J. W., Steel, R. P., Loveland, J. M., & Gibson, L. W. (2004). An investigation of personality traits in relation to adolescent school absenteeism. *Journal of Youth and Adolescence*, *33*(5), 457-466.

MacCann, C., Lipnevich, A., Burrus, J., & Roberts, R. (2012). The best years of our lives? Coping with stress predicts school grades, life satisfaction, and feelings about high school. *Learning and Individual Differences*, *22*(2), 235-241.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*(4), 540-551.

Mahoney, J. L., Durlak, J. A., & Weissberg, R. P. (2018). An update on social and emotional learning outcome research. *Phi Delta Kappan, 100,* 18-23.

Markon, K. E. (2009). Hierarchies in the structure of personality traits. *Social and Personality Psychology Compass*, *3*(5), 812-826.

Markon, K. E., Krueger, R. F., & Watson. D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, *88*(1), 139-157.

Marsh, H. W., & Hau, K. T. (2003). Big-fish--little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364-376.

Martin, J. E., Menten, A., Walton, K. E., Burrus, J., Anguiano-Carrasco, C., Olaru, G., & Roberts, R. D. (2019). *A Rosetta Stone for social and emotional skills: Using the Big Five personality factors to understand, assess, and enhance social and emotional skills in primary and secondary education worldwide.* Iowa City, IA: Asia Society & ACT.

Maydeu-Olivares, A., & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research*, *45*(6), 935-974.

McCloy, R. A., Heggestad, E. D., & Reeve, C. L. (2005). A silk purse from a sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, *8*(2), 222-248.

McCrae, R. R., & Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (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.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730-740.

McDaniel, M. A., & Whetzel, D. L. (2007). Situational judgment tests. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement: Industrial psychology in human resources management* (pp. 235-257). Hillsdale, NJ: Erlbaum.

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*(4), 531-552.

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *Journal of School Health*, *80*(6), 271-279.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574-583.

Organisation for Economic Co-operation and Development. (2015). *Skills for social progress: The power of social and emotional skills*. Paris, France: OECD Publishing.

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

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*(2), 187-207.

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*(3), 524-539.

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

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261-288.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3-25.

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., 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.

Ruus, V., Veisson, M., Leino, M., Ots, L., Pallas, L., Sarv, E.-S., & Veisson, A. (2007). Students' well-being, coping, academic success, and school climate. *Social Behavior & Personality*, *35*(7), 919-936.

Schmitt, D., Allik, J., McCrae, R. R., & Benet-Martinez, 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.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*(6), 1479-1497.

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., & John, O. P. (2014). Traits in transition: The structure of parent-reported personality traits from early childhood to early adulthood. *Journal of Personality*, *82*(3), 182-199.

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science*, *24*(5), 358-362.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, *29*(3), 184-203.

Tackett, J. L. (2006). Evaluating models of the personality-psychopathology relationship in children and adolescents. *Clinical Psychology Review*, *26*(5), 584-599.

Tackett, J. L., Kushner, S. C., De Fruyt, F., & Mervielde, I. (2013). Delineating personality traits in childhood and adolescence: Associations among measures, temperament, and behavioral problems. *Assessment*, *20*(6), 738-751.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385.

Tupes, E. C., & Christal, R. C. (1961). *Recurrent personality factors based on trait ratings*. (USAF A*SD* Technical Report, 61-97). Lackland Air Force Base, TX: U.S. Air Force.

Van Eck, K., Johnson, S. R., Bettencourt, A., & Johnson, S. L. (2017). How school climate relates to chronic absence: A multi-level latent profile analysis. *Journal of School Psychology*, *61*, 89-102.

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

Walton, K. E., Radunzel, J., Moore, R., Burrus, J., Anguiano-Carrasco, C., & Murano, D. (2019). *Adjectives vs. statements in forced choice and Likert item types: Which is more resistant to impression management in personality assessment?* Manuscript submitted for publication.

Wang, L., MacCann, C., Zhuang, X., Liu, O. L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology*, *24*(2), 108-124.

Wang, W., Lee, P., Joo, S., Stark, S., & Louden, R. (2016). MCMC Z-G: An IRT computer program for forced-choice noncognitive measurement. *Applied Psychological Measurement*, *40*(7), 551-553.

Wang, W. C., Qiu, X. L., Chen, C.-W., Ro, S., & Jin, K. Y. (2017). Item response theory models for ipsative tests with multidimensional pairwise comparison items. *Applied Psychological Measurement*, *41*(8), 600-613.

Watson, D. Hubbard, B., & Wiese, D. (2000). General traits of personality and affectivity as predictors of satisfaction in intimate relationships: Evidence from self- and partner-ratings. *Journal of Personality*, *68*(3), 413-419.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, *7*(2), 168-190.

About ACT

ACT is a mission-driven, nonprofit organization dedicated to helping people achieve education and workplace success. Grounded in 60 years of research, ACT is a trusted leader in college and career readiness solutions. Each year, ACT serves millions of students, job seekers, schools, government agencies and employers in the US and around the world with learning resources, assessments, research and credentials designed to help them succeed from elementary school through career. To learn more, visit us at act.org.

ACT®