Article

Journal of Advanced Academics
2020, Vol. 31(4) 470–505

DOI: 10.1177/1932202X20929963
journals.sagepub.com/home/joaa
SAGE

Validation of Instruments for Measuring Affective Outcomes in Gifted Education

Carolyn M. Callahan[1], Amy Price Azano[2], Sunhee Park[1,3], Annalissa V. Brodersen[4], Melanie Caughey[5], Erika L. Bass[6] and Christina M. Amspaugh[1]

Abstract

With increasing attention to examining cognitive strengths and achievements related to social and emotional variables, it is imperative that instruments developed and used to assess change be valid and reliable for measuring underlying constructs. This study examines instruments identified and/or developed to measure four non-cognitive constructs (i.e., student engagement, self-efficacy, growth mindset, and stereotype threat) as outcome variables in a study with elementary-aged students in high-poverty rural communities. The process of creating and examining the psychometric properties of these instruments is a necessary step in documenting the usefulness of the instruments not just in our study but also in other studies with elementary students. We note in our descriptions of the development and

[1]University of Virginia, Charlottesville, USA
[2]Virginia Tech, Blacksburg, USA
[3]Independent Researcher, Palo Alto, CA, USA
[4]Education Consultant, Olathe, KS, USA
[5]Cleveland State University, OH, USA
[6]University of Northern Iowa, Cedar Falls, USA

Corresponding Author:
Carolyn M. Callahan, Department of Curriculum Instruction and Special Education, University of Virginia, 417 Emmet Street, P.O. Box 400277, Charlottesville, VA 22904-4277, USA.
Email: cmc@virginia.edu

assessment of measures that underlying factors may or may not parallel those identified in the general population or in older students and that measurement of noncognitive variables in the population of young gifted students requires considerable attention.

Keywords
assessment, affective outcomes, gifted, validity, reliability

Given the increasing attention to examining the ways gifted students develop and the influence of environments and programs on development, it is imperative that the instruments used to assess developmental or treatment/intervention change be valid and reliable for measuring underlying constructs. These constructs range from constructs relating to cognitive strengths and achievements to constructs relating to social and emotional variables. Educational researchers pay considerable attention to measuring outcomes in studies of programmatic or curricular impact and in program evaluation. Although academic achievement is frequently noted as an outcome of program or curriculum implementation for gifted students (and presents its own set of challenges), many programs and interventions focus on other important outcomes. For example, the recent evolution of theories regarding factors associated with high levels of achievement and the subsequent study of those factors have resulted in the emergence of several other factors, including motivation, mindset, opportunity, creativity, task commitment, interest, and passion, as important in the study of gifted students and interventions for gifted students (Subotnik et al., 2011).

In this article, we discuss four of these noncognitive constructs identified for measurement in a larger study on the impact of implementing a gifted curriculum with elementary school-age gifted children in rural, high-poverty schools: student engagement, self-efficacy, growth mindset, and stereotype threat. Recognizing there are many outcomes that could be part of an efficacy study, we identified constructs related closely to either supporting student achievement or a likely result of successful engagement in an appropriate curriculum and/or instructional setting. Each of these constructs can be considered as impacting the efficacy of a curricular or instructional program or can be a reasonable expected outcome of program intervention. We selected those constructs the literature has identified as particularly relevant to these outcomes and particularly relevant for the historically underrepresented group of gifted students in our project. For example, the degree to which a student is engaged in the lessons of a curriculum or area of study has been identified as a factor in enhancing the impact of curriculum, or, alternatively, a unit of study may serve to increase student engagement in an area of study (Christenson et al., 2012; Fredricks et al., 2004; Jimerson et al., 2003; Reeve & Lee, 2014; Upadyaya & Salmela-Aro, 2013; Wigfield et al., 2015). Self-efficacy and the interrelated

concepts of stereotype threat and growth mindset can similarly be factors that influence the way a student engages in learning, or decreased stereotype threat or increased belief in a growth mindset may be expected outcomes of a treatment (Calkins & Ehrenworth, 2016; Gallagher, 2011; Honicke & Broadbent, 2016; Stewart, 2011). In our study, we focused on these constructs as outcome variables; hence, we needed to identify or create instruments that would be reliable and valid for the data gathering in our study. The process of creating and examining the psychometric properties of these instruments as described in this article is to document the usefulness of the instruments not just in our study but also for other researchers who might wish to assess these constructs in elementary students.

Researchers have developed instruments to measure these constructs in the general school population, but data on the reliability and validity of the instruments for use with gifted populations, and in particular, with gifted populations at the elementary and middle school level, are scant when they exist at all. Because one cannot assume the reliability and validity of the use of data to make decisions in populations where the instruments have not been validated, it follows that these instruments should be examined with samples of those populations. In the following sections, we will first examine the development and prior data on instruments measuring student engagement, self-efficacy, growth mindset, and stereotype threat, particularly with gifted populations when available. Then we will describe ways we developed or modified instruments for use with elementary school-age gifted students and the studies we conducted on the psychometric properties of the revised instruments, beginning with pilot studies and ending with confirmatory studies.

Literature Review

The variables and instruments considered for this study are described not only in terms of the general definition and consideration of the underlying construct but also in terms of their relevancy to the population (gifted students), in the particular domain assessed (language arts), and/or in a particular setting (rural; high-poverty schools) as applicable. This focus on language arts and rural gifted students follows from the overarching aims of the treatment we implemented.

Student Engagement

Definitions of student engagement focus on the intensity and emotional quality of students' involvement in initiating learning activities and their continued involvement in those activities (Connell & Wellborn, 1991; Skinner, 1991). Connell and Wellborn (1991) and Skinner et al. (2009) further distinguished the behavioral and emotional components of engagement. The key markers of engaged behaviors identified by Skinner et al. include the exertion of effort,

persistence, and mental attributes of attention and concentration. These aspects of engagement have been referred to as on-task behavior, academic behavior, and class participation. The emotional attributes of student engagement described by Skinner et al. are reflected in enthusiasm, interest, and enjoyment. Christenson and colleagues expanded the construct to include four subtypes of engagement: academic, behavioral, cognitive, and affective (Appleton et al., 2006; Christenson & Anderson, 2002; Christenson et al., 2008; Reschly & Christenson, 2006).

Relationships between student engagement and student performance. Over the past several decades, classroom engagement has been identified across multiple studies as a factor central to students' academic experience and performance (Christenson et al., 2012; Fredricks et al., 2004; Jimerson et al., 2003; Reeve & Lee, 2014; Upadyaya & Salmela-Aro, 2013; Wigfield et al., 2015). This kind of engagement, which is considered a key marker of academic motivation (Reeve & Lee, 2014), seems to serve both as a gateway to learning (Skinner et al., 2016) and as a protective factor against negative academic outcomes across students' entire educational careers (Finn & Zimmer, 2012). While student engagement has not been studied in gifted populations in general, Landis and Reschly (2013) conclude from a review of the literature that the academic, behavioral, affective, and cognitive components of student engagement are significant themes reflected in the school experiences of gifted students who underachieve and/or drop out.

Measuring student engagement. One measure widely used to assess student engagement is the Engagement vs. Disaffection Scale (EvsD; Skinner et al., 2009), which provides measurement of behavioral and emotional engagement through student self-report, teacher report, and, for a limited number of students, observation (Skinner et al., 2016). The instrument was originally developed for third through sixth graders, and the validity and reliability evidence of the instrument as reported (Skinner et al., 2009) was considered sufficient to justify use of the original version of the instrument. Previous studies (Skinner et al., 2009, 2016) have identified the items from EvsD to best fit a four-factor structure distinguishing Behavioral Disaffection, Behavioral Engagement, Emotional Disaffection, and Emotional Engagement.

Self-Efficacy

Bandura (1997) defined self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). According to Bandura, self-efficacy beliefs can be mitigated by environmental, academic, and personal factors, but are not necessarily based on prior achievement of similar goals. Self-efficacy is embedded within the social structure of the

environment (Bandura et al., 1996); hence, social influences can affect performance. Because self-efficacy is domain-specific, these environmental factors can vary based on the domain under consideration (Pajares et al., 2007). Accordingly, self-efficacy is a factor influencing a person's belief about accomplishing goals in language arts (e.g., reading and writing). If students see themselves as being able to achieve reading and writing goals, they can see themselves as being successful in the language arts classroom (Calkins & Ehrenworth, 2016; Gallagher, 2011; Honicke & Broadbent, 2016; Stewart, 2011).

Self-efficacy of gifted students. Pajares (1996; Pajares & Graham, 1999) provided an example of a measure of self-efficacy as a predictor. In their studies, the researchers found the self-efficacy instrument we adapted for our study predicted problem-solving and overall performance in mathematics. Similarly, Zimmerman and Martinez-Pons (1990) used a measure of self-efficacy to document a relatively small, significant correlation (.56) between math self-efficacy and verbal self-efficacy and between math and verbal self-efficacy and use of self-regulated learning strategies. In addition, Melike (2018) found gifted students' self-efficacy beliefs and sophisticated beliefs about justification and development significantly predict their metacognitive use in science problem-solving. For self-efficacy relating to positive and negative perfectionism, Chan (2007) found that positive and negative perfectionism appeared to directly impact subjective well-being, but self-efficacy played a mediating role between perfectionism and subjective well-being.

Measuring self-efficacy. Because of the emphasis on the curriculum used in our project, our goal was to measure changes in self-efficacy in language arts. Our specific focus was on self-efficacy in writing, reading, and conducting research, which are represented in the standards of the state in which the project was conducted and in the Common Core State Standards in English Language Arts (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). We first identified research on self-efficacy that included instruments to measure self-efficacy in the area of writing (Ainley et al., 2002; Andrade et al., 2009; Pajares, 1996, 2003; Pajares et al., 1999; Shell et al., 1989, 1995). The instruments in these studies were self-report surveys, which participants completed before and after a treatment was conducted. However, these studies did not address self-efficacy in reading or research. Furthermore, existing instruments in the area of writing, designed for older students, were structured with cognitively challenging response categories for children and included items relating to concepts and skills in language arts likely to be unfamiliar to the target age group that included third- and fourth-grade students.

Growth Mindset

The construct of mindsets has gained attention in the last two decades due to the relationship between mindsets and student achievement posited by Dweck and her colleagues (Blackwell et al., 2007; Dweck, 2000; Romeo et al., 2014). According to Dweck (2006), mindset refers to beliefs about one's ability and research has explored how "growth" or "fixed" mindsets affect the learning process. Those with a growth mindset believe ability is malleable and developed through effort and learning; a fixed mindset is rigid and difficult to change (Dweck, 2006). People with a growth mindset believe their ability can be developed, leading them to focus on learning goals and choose challenging tasks even in demanding academic situations. By contrast, when individuals with a fixed mindset face difficult academic situations, they tend to avoid challenging tasks. Because their learning goals focus on performance, they tend to be less willing to choose challenging tasks for fear of seeming incompetent. These two learning processes result in different academic achievement in the face of academic difficulty.

Gifted students and mindsets. Callahan (2012) noted the particular importance of considering and further researching the impact of incremental (growth mindset) versus entity (fixed mindset) theories of intelligence on gifted learners, particularly as one group of researchers (e.g., Mueller & Dweck, 1998) has assumed the "gifted" label might lead gifted students to accept a fixed mindset. Some researchers have found gifted adolescents tend to adopt a growth mindset. For example, Esparza et al. (2014) found gifted seventh-grade students are more likely to endorse a growth mindset in the science domain, and Lüftenegger et al. (2015) found mathematically gifted high-achieving adolescents are less likely to express the belief that intelligence is fixed compared with non-high-achievers. However, Mofield and Peters (2018) found no significant differences between groups on fixed or growth mindset across groups of middle school students identified as gifted, those in advanced class (not identified as gifted), and other students in the school. Research by Makel et al. (2015) suggests gifted students may view the constructs of giftedness and intelligence as independent, with gifted students reporting stronger fixed beliefs about giftedness and more malleable beliefs about intelligence generally.

Measuring mindsets. Most studies assessing mindsets (e.g., Blackwell et al., 2007; Park et al., 2016) have used Dweck's (2000) implicit theories of intelligence scale. This scale comprised eight items on a 6-point Likert-type scale ranging from *strongly agree* to *strongly disagree* with four items to assess a growth mindset. Dweck suggested the 8-item scale can be used for adults and the 6-item scale embedded in the 8-item scale can be used for children older than 10 years old. Researchers (Dweck et al., 1995; Hong et al., 1999; Park et al., 2016)

have evaluated psychometric qualities of the implicit theories of intelligence scale for the general population (Dweck, 2000; Dweck et al., 1995; Hong et al., 1999). These researchers reported strong factorial validity, discriminant validity, and reliability estimates of scores. Recently, Park et al. (2016) reported internal consistency of scores and construct validity evidence of scores from the mindset scale using gifted education populations of middle school and high school age and suggested the 6-item scale can be used for gifted adolescents.

To date, research on mindsets with either the student population in general or the gifted population has focused on measuring mindset in students who are in the upper level of elementary school or beyond. Given that the "gifted" label might lead gifted students to adopt a fixed mindset (Clinkenbeard, 2012; Mueller & Dweck, 1998) and those labels are assigned in many schools as early as second or third grade, it is important for researchers to reliably and validly measure the construct at the lower grade levels and on gifted samples if seeking to understand the developmental onset of mindset adoption of identified as gifted. To date, the instruments used to measure those constructs have not been assessed for their psychometric properties when used with young children nor have they been modified to reflect the language and understandings of young children.

Stereotype Threat

Stereotype threat is a construct often used to explain underachievement and failure to reach full potential (Aronson & Steele, 2005). Stereotype threat is posited as a psychological phenomenon that may hinder the academic achievement of individuals in domains where they perceive negative ability stereotypes about the ethnic, racial, gender, or social group with which they self-associate (Steele, 1997). As a situational threat, stereotype threat can influence the members of any group when they have negative stereotypes about the ability of individuals in their groups relative to other groups (Steele, 1997). When negative perceptions influence beliefs about one's ability, stereotype threat may result in diminished academic performance and lowered career aspirations of individuals who value performance in the stereotyped domains and have the competency to succeed in these areas (Picho & Brown, 2011).

Types of stereotype threat and achievement. Meta-analytic studies on stereotype threat across a wide range of groups (e.g., ethnic minorities, women, older adults, individuals of lower socioeconomic status [SES]) have documented their experiences with stereotype threat (e.g., Nguyen & Ryan, 2008; Picho et al., 2013; Walton & Cohen, 2003; Walton & Spencer, 2009). For example, researchers focusing on issues of race and gender have found Black students and women performed poorly in testing contexts when reminded of their association with stigmatized groups prior to testing (Fiske, 1998; Greenwald & Banaji, 1995;

Spencer et al., 1999; Steele & Aronson, 1995; Wasserberg, 2014). Research on other stereotype threat factors, such as identifying as rural or low SES, also indicates there are negative effects of stereotype threat on academic and affective measures (Désert et al., 2009; Hébert & Beardsley, 2001; Spencer & Castano, 2007; Strayhorn, 2009). Among a group of low-SES students, a group reminded of their socioeconomic identity prior to testing performed more poorly and stated lower self-confidence than similarly low-SES students who were not reminded (Spencer & Castano, 2007). Strayhorn (2009) also revealed that leveled and/or lower aspirations were related to low-SES Black males in both urban and rural areas.

The effects of stereotype threat have also been identified among children (Ambady et al., 2001; Bian et al., 2017; Désert et al., 2009; Wasserberg, 2014). Ambady et al. (2001) have shown the impacts of gender and ethnic stereotypes on the performance of young children. Désert et al. (2009) found that as early as 6 years old, children believed high-SES children perform better at school than those from low-SES backgrounds. Gender stereotypes about intellectual ability emerged as young as six years old and were related to children's interests in the study by Bian et al. (2017).

Stereotype threat and gifted. Stereotype threat has also been demonstrated in high-achieving and/or gifted populations. For example, White male university engineering students performed lower when tested with groups of Asian students. African American students at highly regarded colleges performed lower when told their ability was being measured, and middle school minority students achieved lower scores on aptitude assessments when asked to identify their race prior to assessment; (Aronson et al., 1998, 1999, 2002; Aronson & Inzlicht, 2004; McGlone & Aronson, 2006; Steele & Aronson, 1995, 1998). Findings in the high-achieving and/or gifted populations have been found to mirror findings in the general population across gender, SES, and racial groups.

Measuring stereotype threat. Before Picho and Brown (2011) developed the Social Identities and Attitudes Scale (SIAS), an integrated measure of stereotype threat, stereotype threat was assessed in individual domains such as math identification (Brown & Josephs, 1999), gender identification (Luhtanen & Crocker, 1992), and stigma consciousness (Brown & Pinel, 2003). Picho and Brown (2011) developed and integrated instrument, SIAS, in response to the notion that the general construct of stereotype threat should be and could be assessed using an integrated measure of stereotype threat. Reliability estimates for the six factors of the SIAS (mathematics, gender, ethnicity, stigma consciousness, and negative affect) ranged from .81 to .95, and bivariate subscale correlations provided support for convergent and discriminant validity. These reliabilities compared favorably with reported reliability estimates of subscales in the studies using isolated stereotype threat. The authors of the SIAS provided evidence of

reliability and validity for the instrument as a measure of stereotype threat factors of the SIAS only for college-age students, and the SIAS does not include assessment of stereotype threat for younger students or as it manifests in rural settings or in low-SES or the language arts domain.

Developing and/or Revising and Piloting the Instruments

Given the status of the instruments described above and the need for reliable and valid instruments for use with gifted children at the elementary school level through the middle school level, we developed or revised two instruments—one for stereotype threat and one for self-efficacy. We then gathered data on those instruments to assess reliability and validity. We also examined the psychometric properties of two instruments that did not require major revision (student engagement and mindset) in the elementary to middle school-age population of third- through eighth-grade gifted students and expanded on the particular constructs described in the instruments above to include stereotype threats of being from rural areas and/or low income.

The samples used for pilot and confirmatory analyses came from several programs. While the school districts included had varying particular definitions of gifted and identification procedures, they could all be categorized under the talent development paradigm as described by Dai and Chen (2013). That is in each case the identification process, while sometimes including a measure of aptitude, was designed to designate learners from a broader, more diverse range of strengths and interests with the intent of fostering excellence in the area of talent in which the student was identified in chosen areas.

Revisions of Scales

Student engagement. To measure student engagement, we examined the properties of a student self-report measure of both behavioral and emotional engagement versus disaffection (Skinner et al., 2009). Because the scale was originally validated on a population similar in age to the population in the current study (third through sixth graders), it was piloted with only minor revisions of several items and elimination of two items that appeared inappropriate or unclear as written.

Self-efficacy. The lack of appropriate assessment tools to measure self-efficacy in the specific academic areas led us to the development of the "How I Feel About Writing and Reading" (Self-Efficacy) scale to measure self-efficacy in writing, reading, and doing research for mid to upper elementary-age students. We first delineated definitions for self-efficacy in writing and reading.

Reading self-efficacy: Beliefs about one's ability to learn and execute skills in comprehending a written text;
Writing self-efficacy: Beliefs about one's ability to learn and execute skills in expressing information persuasively and effectively in a written form;
Research self-efficacy: Beliefs about one's ability to learn and execute the skills needed to investigate and report from varied sources on a new or an unknown topic with accountability.

Next, we referenced the English content standards in the state in which the study was conducted and general language arts standards as referenced in the Common Core State Standards for Language Arts (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) to ensure lessons developed in the curriculum related to reading, writing, and research aligned with grade-level expectations.

We selected and adapted items of the self-efficacy instrument from Pajares et al. (1999) and added additional items pertinent to the domains described above. The preliminary scale included more than 40 items, including revised items from the existing instrument and the newly constructed items reflecting self-efficacy in reading and writing. Experts in language arts with experience teaching third and fourth graders and experts in gifted education with experience developing instruments targeted for third and fourth graders identified as gifted reviewed the items for content and language structure, and changes were made accordingly. We reduced the scale to 15 by eliminating redundant items and preserving items reflecting key skills in language arts as identified by the language arts experts on our team. All items were phrased with the "can do" language for self-efficacy scales as recommended by Bandura (2006). Although Bandura (2006) recommended a response category with a range of 0 to 100, we considered the age of the respondents and created four-response categories with detailed descriptions for each choice (Bell, 2007).

Because of the language and cognitive complexity of the constructed and revised items, two research team members conducted nine cognitive interviews with third-grade students identified either as gifted or as advanced readers from a rural school district (target group). The cognitive interviews provided us an opportunity to clarify language used in the instrument based on insights on how students interpreted the items (Bell, 2007). The interview questions focused on (a) the overall structure of the scale, (b) clarification of vocabulary choices (e.g., grammar) in the scale, (c) the ease at which the students could complete the scales (e.g., in relation to response choices), and (d) the instructions provided to complete the scale. Interviewers provided written reports on their cognitive interviews. Minor alterations and clarifications to the scale resulted in a 15-item scale.

Mindset. To measure implicit theories of intelligence, we selected the Implicit Theories of Intelligence Scale (Dweck, 2000). Dweck and her colleagues have

offered two different measures of implicit intelligence: a six-item scale and an eight-item scale (Dweck, 2000). Dweck (2000) has recommended the 6-item scale for use with children aged 10 years or older and the eight-item scale for use with adult populations. However, those researchers had not presented sufficient validity evidence for the structure of the six-item scale in children younger than 10 years old, nor have any researchers provided validity and reliability evidence for use of the scores of the six-item scale with gifted elementary students. Because our target population was elementary students, we selected the six-item scale. Once again, we conducted cognitive interviews with nine third- and fourth-grade students with the same focus noted above to assess the readability and clarity of items. We revised items as needed and renamed the scale "How Does Your Brain Work?".

Stereotype threat. Although traditional research in the area of stereotype threat has focused on issues of race and gender (e.g., Fiske, 1998; Greenwald & Banaji, 1995), researchers recently have studied stereotype threat factors such as rural settings or low SES (Hébert & Beardsley, 2001; Spencer & Castano, 2007). In addition, although early research focused on stereotype threat in students in middle school or older, Ambady et al. (2001) have shown the impacts of stereotypes related to gender identity and ethnic identity on the performance of young children. In our study, we investigated stereotype threat relating to gender, ethnicity, rural settings, low-SES status, and reading/writing domains for elementary students living in impoverished rural domains.

To create the "Who I Am and How I Learn" (WIAHIL) scale measuring various stereotype threat factors, we first examined the integrated measure of stereotype threat, the SIAS (Picho & Brown, 2011). We selected gender, ethnicity, and stigma consciousness factors from the SIAS (Picho & Brown, 2011) and revised the items from those factors to reflect the reading level and language of elementary school students. We then identified four additional factors we considered important potential factors influencing stereotype threat relevant to the project focus and the setting. We hypothesized that beliefs about living in rural settings and in poverty would be stereotype threat factors for students living in economically distressed rural domains. We also posited that beliefs about gender competence in reading/writing might affect performance in a language arts unit and on language arts assessments. We created new items representing the new factors based on the SIAS approach and format. The resulting instrument, WIAHIL, included additional items designed to measure stereotype threat related to status as rural students and relative to their status as students of low SES and in relation to stereotype threat in the domains of reading/writing. Experts in gifted education, literacy, and elementary education reviewed the items in the new WIAHIL scale. Based on the reviews, we divided the survey into two forms: one for girls and the other for boys (to reduce the complexity of language and

cognitive load created by using girl/boy in each stem relating to gender). We again conducted cognitive interviews with nine third and fourth graders identified as gifted in their school districts and further revised items based on their responses to individual items and the scale as a whole. This process resulted in a 45-item scale. While a 45-item scale was deemed likely too long for elementary school students, we determined it would be best to eliminate items based on their psychometric properties.

Pilot Tests[1]

Student engagement. To pilot test the 45-item engagement scale, we recruited a sample of 191 fifth- through eighth-grade students participating in a university-based enrichment program[2] (females = 49.7%). We conducted an exploratory factor analysis (EFA) based on a polychoric correlation matrix (due to the use of a four-point rating scale) to assess whether the items reflect the proposed substructure of the scale in a gifted sample. Although the authors of the original scale identified four subfactors (Behavioral Engagement, Emotional Engagement, Behavioral Disaffection, and Emotional Disaffection), eigenvalues greater than 1 indicated six subfactors. All fit indices of a four-factor model were in either an acceptable or a good range (root mean square error of approximation [RMSEA] = .07; comparative fit index [CFI] = .97; Tucker–Lewis index [TLI] = .93; standardized root mean residual [SRMR] = .06), and factor loadings of a four-factor model were clearer than factor loadings of the fixed or six-factor model. Thus, we followed up with an EFA with an oblique rotation for a four-factor model. Based on these results of the factor loadings on the EFA, six items were removed, leaving three factors (Behavioral Engagement, Emotional Engagement, and Disaffection).

After removing the six items, we conducted a second EFA. Model fit indices of the three-factor model were in either an acceptable (CFI = .95; TLI = .93; SRMR = .07) or an inadequate (RMSEA = .09) range, whereas all model fit indices of the four-factor model were in a good range (CFI = .97; TLI = .95; SRMR = .05; RMSEA = .07). Also, eigenvalues greater than 1 indicated four factors. An EFA with an oblique rotation was performed for the four-factor model. In the second EFA, the results were similar to the results in the first EFA. Five items loaded on the first factor (Behavioral Engagement), and four items loaded on the second factor (Emotional Engagement). Six items loaded on the third factor, and two items loaded on the fourth factor. One item cross-loaded on two factors, and one item cross-loaded on three factors.

Based on a factor structure and factor loadings in the second EFA (Table 1), we selected Items 1, 2, 3, 4, and 5 to create a Behavioral Engagement factor and Items 6, 7, 8, and 9 for an Emotional Engagement factor. Factor loadings of those nine items were all above 0.40 except for Item 3 (0.37). Reliability estimates (internal consistency assessed by polychoric ordinal alpha) were .85 for

Table 1. Student Engagement: Factor Structure and Factor Loadings From the Second EFA With 19 Items (Pilot Analysis).

	Rotated factor structure			
Item	1	2	3	4
1. I try hard to do well in school.	.83			
2. In class, I work as hard as I can.	.77			
3. When I'm in class, I participate in class discussions.	.37			
4. I pay attention in class.	1.09[a]			
5. When I'm in class, I listen very carefully.	.69			
6. When I'm in class, I feel good.		.57		
7. When we work on something in class I feel interested.		.70		
8. Class is fun.		.95		
9. I enjoy learning new things in class.		.74		
12. I don't try very hard in school.	–.59		.35	
15. When we work on something in class, I feel bored.	–.29	–.33	.33	
16. When I'm in class, I feel worried.			.97	
16a. When we start something new in class, I feel worried.			.82	
16b. When I get stuck on a problem, I feel worried.			.46	
17. When we work on something in class, I feel discouraged.			.80	
19. When I'm in class, I feel bad.			.72	
19a. When I'm working on classwork, I feel mad.			.66	
19b. When I get stuck on a problem, it really bothers me.				.93
19c. When I can't answer a question, I feel frustrated.				.86

Note. EFA = exploratory factor analysis.
[a]Standardized values above 1.0 can sometimes be valid (Jöreskog, 1999), and in this case, estimated residual variance of this item was positive.

the Behavioral Engagement factor and .87 for the Emotional Engagement factor. The other items either cross-loaded or did not form coherent factors.

Self-efficacy. In the first stages of validation, we conducted two investigations on the items we had selected/adapted or added to the scale from Pajares et al. (1999). Based on descriptive findings from the first investigation (using 15 items as described above) with a sample of 42 students in a summer program for gifted students, we revised items and factors and reassigned items to new,

logically connected factors. This sample was too small for factor analysis, but patterns of response indicated some items did not associate with other logically related items. As a result, 12 items indicating three factors (self-efficacy for completing simple writing and reading task, self-efficacy for completing complex writing and reading task, and self-efficacy in conducting research) were included in the instrument used for the second step of validation.

For the second investigation, using the 12-item survey, we recruited a sample of 191 students from those participating in a summer program for gifted students (females = 49.7%). We conducted an EFA using a principal axis factoring (PAF) with oblique rotation (oblimin) to determine whether the items reflected the proposed substructure of the scale in gifted populations. The Kaiser–Meyer–Olkin (KMO = .849) measure of sampling adequacy and Bartlett's test of sphericity ($\chi^2 = 822.356$, $df = 55$, $p < .000$) were tested before running the EFA. Eigenvalues greater than 1 and the scree plot from the EFA indicated three factors. Horn's parallel analysis (Horn, 1965) suggested two factors. However, in the two-factor solution, the second factor contained only three items and the first factor contained the rest of the items. Hence, based on these two analyses (two- and three-factor solutions), the three-factor solution was retained. The total variance explained by those three factors was 50%. The results of an EFA with an oblique rotation (oblimin) resulted in the loading of four items from the original Complex scale and one item from the original Simple factor on a first factor. Two items from the Simple scale and two items from the Research scale loaded on a second factor. Three items (one from the Simple scale, one from the Complex scale, and one from the Research scale) loaded on a third factor. Eleven factor loadings were in an acceptable range which is above .40 (Hair et al., 1998); however, the loading for one item was considerably less than .40 (.27).

Thus, we reconducted an EFA after removing the item that failed to load on any factor. Again, the scree plot and eigenvalues from the second EFA identified three factors, whereas Horn's parallel analysis (Horn, 1965) suggested two factors. In the two-factor solution, the second factor contained only three items and the first factor contained the rest of the items. For this reason, a three-factor solution was retained. Total variance explained by the three factors identified was 52%. We then conducted an EFA with an oblique rotation (oblimin). As shown in Table 2, Items 1, 2, 4, 5, and 6 loaded on the first factor. Items 7, 8, and 12 loaded on the second factor, and Items 3, 9, and 10 loaded on the third factor. However, examination of the content of the items in this analysis did not suggest a logical grouping of items across factors. Furthermore, the reliability estimate (internal consistency) across all 11 items was .87 (Cronbach's alpha). Thus, we concluded, due to the overall high internal consistency and lack of logical interpretation of the factor analytic results, that the instrument represents one single general scale measuring self-efficacy in reading and writing, without any subscales.

Table 2. Self-Efficacy: Factor Structure and Factor Loadings From the Second EFA With 11 Items (Pilot Analysis).

	Rotated factor structure		
Item	1	2	3
2. I use clues from a story or poem to understand how the author wants me to feel when I am reading it.	**.72**	–.09	–.04
6. I can correct my own work to make sure anyone who reads it will understand what I'm trying to say.	**.68**	.16	–.01
4. I can write a good story that a reader can easily understand.	**.60**	–.05	.20
1. I can use clues from a story or poem to understand what the author is trying to say.	**.53**	.05	.26
5. I can correctly use the writing strategies that I have learned in class.	**.56**	.32	–.10
7. I can correctly spell all the words in a one-page story.	.11	**.68**	–.07
8. I can write a simple sentence using correct rules for writing.	.01	**.64**	.30
12. I know how to show that information in my writing has come from difference sources, such as the internet, books, magazines, or people.	.14	**.52**	.16
10. I can find information that I need from more than one source, such as the internet, books, magazines, or people.	.01	.09	**.75**
9. I can ask good questions that will help me find out more about a topic I am interested in.	.34	–.25	**.50**
3. I can read smoothly and easily.	.07	.21	**.60**

Note. EFA = exploratory factor analysis. Loadings in bold indicate high loadings on the factor of that column.

Mindset. Because of the extensive analysis of the mindset scale in prior studies, we did not conduct a pilot assessment (e.g., Blackwell et al., 2007; Dweck et al., 1995; Haimovitz & Dweck, 2016; Park et al., 2016).

Stereotype threat. For the first pilot, we administered the scale to two samples. The first sample comprised third- and fourth-grade students who participated in a Saturday university-based enrichment program for gifted students. The second sample included gifted students identified in third and fourth grade in a local school district. For the second sample, at the request of the local school district, we administered the survey without items designed to measure ethnic

identification and ethnic stigma consciousness factors. Of the 101 participants from the two samples, 85 completed the survey. Due to the small sample size, we conducted a descriptive data analysis and calculated internal consistency estimates of the initially proposed factors which ranged from .59 (low-SES identification) to .89 (reading/writing identification).

Also, due to the small sample size, we conducted a PAF analysis with oblique rotation (oblimin) including only the newly created items measuring Rural Background Identification, Rural Background Stigma Consciousness, Low-SES Identification, Low-SES Stigma Consciousness, and Reading/Writing Identification factors. Items measuring gender identification and gender stigma consciousness factors were excluded for the EFA because those items were not newly created ones, and items measuring ethnic identification and ethnic stigma consciousness factors were also excluded because we were unable to include those items in the assessment of the second sample.

The KMO (.72) measure of sampling adequacy and Bartlett's test of sphericity ($\chi^2 = 1{,}177.59$, $df = 351$, $p < .000$) were computed before running the PAF. The scree plot and eigenvalues greater than 1 indicated seven factors, whereas Horn's parallel analysis (Horn, 1965) suggested five factors. In the seven-factor solution, the sixth factor contained only one item. Because we initially included items from hypothesized five factors (Rural Background Identification, Rural Background Stigma Consciousness, Low-SES Identification, Low-SES Stigma Consciousness, and Reading/Writing Identification factors) and this was supported by Horn's parallel analysis, we decided to retain a five-factor solution as the parallel analysis suggested.

The total variance explained by the identified five factors was 54.5%. All communalities were in an acceptable range above .40 except for two items (.19 and .16). Five items from the proposed Low-SES Stigma Consciousness factor and three items from the proposed Rural Background Stigma Consciousness factor again loaded on a second factor. The Proposed Reading/Writing Identification items loaded on two separate factors, with five items each loading on the first and second factors. Five items from the proposed Rural Background Identification Factor and one item from the Proposed Rural Background Stigma Consciousness factor loaded on a third factor. Two items from the proposed Low-SES Identification Factor loaded on a fifth factor, whereas only one (from the proposed Low-SES Identification factor) loaded on the last factor. Based on the results of the PAF with an oblique (oblimin) rotation and the five-factor solution, items with communalities less than .4 were removed from the scale. One item was also removed because it cross-loaded on two factors. Because the instrument was still very long, we also removed four items because they appeared to be redundant and not clearly representative of stereotype threat. Also, one item from the proposed Low-SES Identification factor was revised because this item loaded on a factor other than the emergent Low-SES Identification Factor.

With these modifications, 38 items were retained for the scale now comprising five newly created factors, including an Integrated factor (Low-SES and Rural Background Stigma Consciousness), a Reading Identification factor, a Writing Identification factor, a Rural Background Identification factor, and a Low-SES Identification factor and the four other factors adapted from the SIAS: Gender Identification, Gender Stigma Consciousness, Ethnic Identification, and Ethnic Stigma Consciousness. As we were unable to include the four factors from the original SIAS (Gender Identification, Gender Stigma Consciousness, Ethnic Identification, and Ethnic Stigma Consciousness) in this pilot test due to the small sample, we needed to conduct an EFA including all nine factors.

We conducted a second pilot test of the stereotype threat scale because of the limited number of subjects in the first pilot study. We recruited our sample from students participating in a summer enrichment program. One hundred ninety-one fifth- through eighth-grade students agreed to participate (females = 49.7%). PAF analysis with oblique rotation (oblimin) was used to determine whether the items reflect the proposed nine-factor structure of the scale. The KMO measure of sampling adequacy had a value of .82, and Bartlett's test was statistically significant ($\chi^2 = 3{,}526.463$, $df = 703$, $p < .000$). The scree plot and eigenvalues greater than 1 indicated nine factors, and parallel analysis (Horn, 1965) suggested six factors. In the six-factor solution, the total variance explained by the six factors was 52%, which was less than the total variance of the nine factors, and examination of the content of some items in the six-factor solution did not suggest a logical grouping of items across factors. Thus, a nine-factor solution was retained. The total variance explained by the nine factors identified was 58.82%. As shown in Table 3, items from the Low-SES Stigma Consciousness scale (26, 30, 31, 34, 35) and items from the Rural Background Stigma Consciousness scale (23, 33, 37) loaded on the first factor. This result indicated those items (23, 26, 30, 31, 33, 34, 35, 37) measure the same construct as we found in the first pilot test. Item 1 from the Gender Identification scale and items from the Gender Stigma Consciousness scale (3, 8, 11, 16, 19) loaded on a fourth factor. Thus, we decided to move Item 1 to the Gender Stigma Consciousness scale. Items from the Gender Identification scale loaded on two separate factors: Items 14 and 2 loaded on an eighth factor and Items 6 and 9 loaded on a ninth factor. Item 29 from the Low-SES Identification scale also loaded on the ninth factor. Item 12 from the Ethnic Identification items and items from the Ethnic Stigma Consciousness (5, 15, 17, 21) loaded on a sixth factor.

Reading Identification items (27, 32, 36) and Writing Identification items (24, 28, 38) loaded together on the second factor. Gender Stigma Consciousness items (3, 8, 11, 16, 19), Ethnic Stigma Consciousness items (5, 15, 17, 21), Ethnic Identification items (4, 10, 13), Rural Identification items (7, 18, 20), and Low-SES Identification items (22, 25) loaded on the factors as expected. All factor loadings were in an acceptable range which is above .40

Table 3. Stereotype Threat: Factor Structure and Factor Loadings After Oblique Rotation (Oblimin) Rotation (Pilot 2).

	Rotated factor structure								
Item	1	2	3	4	5	6	7	8	9
31. Other people judge me because of how much money my family has.	.78								
35. People think I act certain ways because of how much money my family has.	.77								
37. Other people judge me because of the community where I live.	.71								
33. My teachers think I act certain ways because of the community where I live.	.59								
23. People think I act a certain way because of the community where I live.	.59								
34. How much money my family has is a reason I act the way I do around other people.	.57								
30. How much money my family has affects how people my age act toward me.	.55								
26. How much money my family has affects how teachers act toward me.	.50								
28. Being good at writing will be useful to me after I finish school in my future.		.86							
27. Being good at reading will be useful to me after I finish school in my future.		.84	.						
36. Being a good reader matters to me.		.83							
32. Being a good reader is important to my success in school.		.83							
38. Being a good writer is important to my success in school.		.77							
24. Being a good writer matters to me.		.60							
10. I would rather have the color of my skin than any other color.			.79						
4. I really like the color of my skin.			.61						
13. The color of my skin is very important to me.			.51						
8. Being a boy (or a girl)[a] affects how other people think about my behavior.				–.76					
16. Girls (or boys) think I act certain ways because I am a boy (or a girl).				–.73					
3. My teacher thinks I act the way I do because I am a boy (or a girl).				–.65					
				–.63					

(continued)

Table 3. Continued

Item	Rotated factor structure								
	1	2	3	4	5	6	7	8	9
11. My teacher treats me differently because I am a boy (or a girl).									
19. Other people treat me differently because I am a boy (or a girl).				–.61					
1. Being a boy (or a girl) influences how I feel about myself.				–.39					
20. My community is very important to me.					.72				
18. I would rather live here than anywhere else.					.47				
7. I know many things about my community.					.41				
21. People judge how I act based on the color of my skin.						–.70			
5. People judge me on the basis of the color of my skin.						–.64			
15. Students in my class treat me differently because of the color of my skin.						–.58			
12. The color of my skin is an important part of who I am.						–.52			
17. Teachers act differently toward me in class because of the color of my skin.						–.50			
22. The things my family has influence how I feel about myself.							–.79		
25. The things my family has influence how confident I am in myself.							–.62		
14. Being a boy (or a girl) makes me a better writer.								.85	
2. Being a boy (or a girl) makes me a better reader.								.85	
6. Being a boy (or a girl) helps people know who I am.									–.61
9. Being a boy (or a girl) is important to me in knowing who I am and how I should act.									–.43
29. Where I live and what I have influence what I can become in the future.									–.28

Note. Rotated factor structure: 1 = Low-SES and Rural Background Stigma Consciousness; 2 = Reading and Writing Identification; 3 = Ethnic Identification; 4 = Gender Stigma Consciousness; 5 = Rural Identification; 6 = Ethnic Stigma Consciousness; 7 = Low-SES Identification; 8 = Gender Identification; 9 = Gender Identification.

[a]We have inserted the alternative gender notation. In administering the instrument, we actually administered two separate forms based on recommendations from experts on reducing the cognitive complexity of the instrument.

(Hair et al., 1998), except for Item 29, which loaded at −.28, and Item 1, which loaded at −.39. Item 29 from the Low-SES Identification scale was revised because the item loaded on a seventh factor, Gender Identification.

Finally, the scale comprised 38 items with eight factors called Gender Stigma Consciousness, Ethnic Identification, Ethnic Stigma Consciousness, Rural Identification, Low-SES Identification, Reading/Writing Identification factors, a single Gender Identification factor which combined the two Gender Identification factors, and Rural Background and Low-SES Stigma Consciousness (see Table 3).

Reliability estimates of all factors proposed from an EFA (internal consistency assessed by Cronbach's alpha) ranged from .52 (Rural Background Identification) to .90 (Rural Background and Low-SES Stigma Consciousness). The reliabilities of all eight factors are summarized in Table 4.

Final Evaluation of the Assessment Tools

Following pilot testing and creation of the final versions of the instruments, we conducted a final round of data collection to examine the psychometric properties of the revised instruments. The sample for the final factor analyses included third-grade students identified as gifted in our project and a new sample of rising fifth- through eighth-grade students identified as gifted for attendance at a summer enrichment program.

Data Analysis

Confirmatory factory analysis (CFA) was performed using Mplus 7 program on all four instruments. To determine the scales with the best fit using the CFAs, we compared approximate fit indices including RMSEA, CFI, and SRMR.

Table 4. Stereotype Threat: Estimates of Internal Consistency (Pilot 2).

Subfactors	Cronbach's α
Ethnic Identification (EI)	.70
Gender Identification (GI)	.69
Gender Stigma Consciousness (GSC)	.83
Ethnic Stigma Consciousness (ESC)	.83
Reading and Writing Identification (R/WI)	.89
Rural Background Identification (RI)	.52
Low-SES Identification (LSI)	.62
Rural Background and Low-SES Stigma Consciousness (RSC)	.90

The criteria for acceptable model fit are less than .08 for both RMSEA and SRMR and greater than .90 for CFI (Bentler, 1990; Browne & Cudeck, 1993; Hu & Bentler, 1999).

Self-Efficacy

A CFA on the self-efficacy based on the results of the pilot data indicating a single factor on the scale for measuring self-efficacy across reading and writing further affirmed the best model fit is for one general factor (see Table 5). All of the indices (SRMR, RMSEA, and CFI) were in an acceptable range. The reliability estimate (internal consistency) was .89 (Cronbach's alpha).

Student Engagement

A CFA based on the results of the pilot studies resulted in a CFI in the good range and the RMSEA fell in an acceptable range. For this model, there were two factors Behavioral Engagement and Emotional Engagement as predicted from the pilot studies (see Table 6). The reliabilities of these factors are .84 and .85, respectively.

Growth Mindset

The CFA for the Growth Mindset scale based on the two factors of Entity and Incremental Theories of Intelligence resulted in both CFI and SRMR indices in the good range, but the RMSEA was not in an acceptable range (see Table 7). Modifications of the model based on the highest modification index resulted in a standardized estimate of the residual correlation of 999, meaning residual variance of the indicator is negative and the residual correlation was not computed. In this situation, we can use normalized residual correlation, which was .004. Negative residual variance signifies model misspecification, skewed data, or small sample size. Our sample size was large, and the data were not skewed. We concluded the modified model is not appropriate even though RMSEA became an acceptable value. The reliability estimates for the fixed mindset and the growth mindset were .85 and .73, respectively.

Stereotype Threat

Four-factor analyses were run on the data from 940 students. The first analysis was a CFA based on the underlying hypothesized dimensions from initial construction of the instrument. The resulting SRMR and RMSEA fell in a range considered acceptable, but the CFI for the scale was not in an acceptable range. Subsequently, Model 1 was modified based on the modification indices, and a second CFA was computed. Based on the modification indices of Model 1, two residual correlations were added to improve the overall model fit. In Model 2, the CFI fell in an acceptable range, and the SRMR and the RMSEA are both in

Table 5. Self-Efficacy: Standardized Estimates of CFA ($N = 936$) on the Final Sample.

Factors	Items	Estimate	SE	RV	V
SE	1. I can use clues from a story or poem to understand what the author is trying to say.	.707	.019	.500	1.000
	2. I use clues from a story or poem to understand how the author wants me to feel when I am reading it.	.693	.019	.520	
	3. I can read smoothly and easily.	.541	.025	.708	
	4. I can write a good story that a reader can easily understand.	.590	.024	.652	
	5. I can correctly use the writing strategies that I have learned in class.	.694	.019	.519	
	6. I can correct my own work to make sure anyone who reads it will understand what I'm trying to say.	.638	.022	.593	
	7. I can correctly spell all the words in a one-page story.	.573	.024	.671	
	8. I can write a simple sentence using correct rules for writing.	.696	.019	.515	
	9. I can ask good questions that will help me find out more about a topic I am interested in.	.670	.021	.551	
	10. I can find information that I need from more than one source, such as the internet, books, magazines, or people.	.677	.020	.542	
	11. I know how to show that information in my writing has come from difference sources, such as the internet, books, magazines, or people.	.688	.020	.527	

Fit statistics	
$\chi^2(df, p)$	245.392 (44, <.001)
CFI	.947
SRMR	.035
RMSEA [90% CI]	.070 [.062, .079]

Note. CFA = confirmatory factor analysis; RV = residual variance; V = variance; SE = standard error; CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval. The final items included in the self-efficacy scale are found in Supplemental Appendix A.

Table 6. Student Engagement: Standardized Estimates of CFA ($N = 934$) on the Final Sample.

Factors	Items	Estimate	SE	V	Factor correlation
BEG	1. I try hard to do well in school.	.754	.033	1.000	.633
	2. In class, I work as hard as I can.	.818	.021		
	3. When I'm in class, I participate in class discussions.	.514	.033		
	4. I pay attention in class.	.851	.021		
	5. When I'm in class, I listen very carefully.	.808	.021		
EEG	6. When I'm in class, I feel good.	.790	.022	1.000	
	7. When we work on something in class I feel interested.	.694	.024		
	8. Class is fun.	.814	.021		
	9. I enjoy learning new things in class.	.803	.022		
	Fit statistics				
$\chi^2(df, p)$	155.307 (26, <.001)				
CFI	.974				
WRMR	1.303				
RMSEA [90% CI]	.073 [.062, .084]				

Note. CFA = confirmatory factor analysis; SE = standard error; V = variance; BEG = behavioral engagement; EEG = emotional engagement; CFI = comparative fit index; WRMR = weighted root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval. The final items for the student engagement scale are found in Supplemental Appendix B.

an acceptable range. A third CFA was conducted based on the results of the EFA. The resulting model yielded an SRMR and RMSEA in an acceptable range, but the CFI for the scale was not in the acceptable range. Model 3 was modified based on the modification indices with six residual correlations added to improve the overall model fit. The SRMR and RMSEA for Model 4 were found to be in an acceptable range, but the CFI for the scale was not in an acceptable range. Based on an examination of all of the models, we concluded Model 2 should be considered the best fit for the data and that it reflected the best assessment of the underlying structure having these factors: Gender Identification, Gender Stigma Consciousness, Ethnicity Identification, Ethnicity Stigma Consciousness, Rural Identification, Rural Stigma Consciousness, Low Income Identification, Low Income Stigma Consciousness, and Reading and Writing Identification. The analysis results of Model 2 are found in Table 8. The reliabilities of the factors ranged from .50 to .85. Table 9 contains all factor reliabilities.

Table 7. Mindset: Standardized Estimates of CFA ($N = 933$).

Factors	Items	Estimate	SE	RV	V	Factor correlation
ET	1. You have a certain amount of intelligence (how smart you are), and you can't really do much to change it.	.814	.016	.337	1.000	–.348
	2. How smart you are is something that you can't change very much.	.922	.014	.149		
	5. You can learn new things, but you can't really change how smart you are.	.697	.020	.665		
IT	3. No matter who you are, you can make a big difference in how smart you are.	.579	.030	.277	1.000	
	4. You can always greatly change how smart you are.	.850	.030	.514		
	6. No matter how smart you are, you can always change it quite bit.	.633	.030	.599		

Fit statistics	
$\chi^2(df, p)$	73.144 (8, <0.001)
CFI	.968
SRMR	.047
RMSEA [90% CI]	.093 [.075, .114]

Note. CFA = confirmatory factor analysis; SE = standard error; RV = residual variance; V = variance; ET = entity theory of intelligence; IT = incremental theory of intelligence; CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval. The final items for the growth mindset scale are found in Supplemental Appendix C.

Discussion

Movement in the field of gifted education from only assessing cognitive strengths to considering more affective domains is important, especially when considering vulnerable populations. Thinking about social and emotional variables is a worthwhile endeavor, but assessing change in those domains and attributing that change to programmatic or curricular interventions should be done with care. In our work, we explicitly focused on rural students in high-poverty communities as an underrepresented population in the literature on gifted education. However, we recognize rural students may be particularly

Table 8. Stereotype Threat: Standardized Estimates of CFA ($N = 940$) and Best Model Fit and Factor Correlations for the Final Sample.

Factors	Items	Estimate	SE	RV	V	Factor correlation
GI	1. Being a boy influences how I feel about myself.	.543	.030	.705	1.000	
	2. Being a boy makes me a better reader.	.520	.031	.730		
	6. Being a boy helps people know who I am.	.643	.027	.586		
	9. Being a boy is important to me in knowing who I am and how I should act.	.614	.028	.623		
	14. Being a boy makes me a better writer.	.438	.033	.808		
GSC	3. My teacher thinks I act the way I do because I am a boy.	.590	.025	.652	1.000	
	8. Being a boy affects how other people think about my behavior.	.705	.021	.504		
	11. My teacher treats me differently because I am a boy.	.695	.021	.517		
	16. Girls think I act certain ways because I am a boy.	.657	.023	.569		
	19. Other people treat me differently because I am a boy.	.727	.020	.472		
EI	4. I really like the color of my skin.	.513	.028	.737	1.000	
	10. I would rather have the color of my skin than any other color.	.498	.028	.752		
	12. The color of my skin is an important part of who I am.	.775	.018	.399		
	13. The color of my skin is very important to me.	.878	.016	.230		
ESC	5. People judge me on the basis of the color of my skin.	.704	.020	.505	1.000	
	15. Students in my class treat me differently because of the color of my skin.	.796	.016	.366		
	17. Teachers act differently toward me in class because of the color of my skin.	.708	.019	.498		
	21. People judge how I act based on the color of my skin.	.800	.016	.361		
RI	7. I know many things about my community.	.462	.035	.786	1.000	

(continued)

Table 8. Continued

Factors	Items	Estimate	SE	RV	V	Factor correlation
	18. I would rather live here than anywhere else.	.388	.036	.849		
	20. My community is very important to me.	.790	.040	.376		
RSC	23. People think I act a certain way because of the community where I live.	.662	.023	.561	1.000	
	33. My teachers think I act certain ways because of the community where I live.	.678	.022	.540		
	37. Other people judge me because of the community where I live.	.833	.016	.307		
LI	25. The things my family has influence how confident I am in myself.	.728	.026	.470	1.000	
	29. Where my family lives and the things my family has influence what I can become in the future.	.540	.031	.709		
	22. The things my family has influence how I feel about myself.	.715	.026	.489		
LSC	26. How much money my family has affects how teachers act toward me.	.663	.021	.560	1.000	
	30. How much money my family has affects how people my age act toward me.	.746	.018	.444		
	31. Other people judge me because of how much money my family has.	.781	.016	.390		
	34. How much money my family has is a reason I act the way I do around other people.	.649	.022	.579		
	35. People think I act certain ways because of how much money my family has.	.774	.016	.401		
RWI	24. Being a good writer matters to me.	.585	.025	.658	1.000	
	27. Being good at reading will be useful to me after I finish school in my future.	.732	.019	.464		
	28. Being good at writing will be useful to me after I finish school in my future.	.724	.019	.476		

(continued)

Table 8. Continued

Factors	Items	Estimate	SE	RV	V	Factor correlation
	32. Being a good reader is important to my success in school.	.770	.017	.407		
	36. Being a good reader matters to me.	.721	.019	.480		
	38. Being a good writer is important to my success in school.	.752	.018	.434		
	GI14 with GI2 EI10 with EI4	.539	.026			
GSC with	GI					.393
EI with	GI					.516
	GSC					.084
ESC with	GI					.207
	GSC					.674
	EI					.168
RI with	GI					.277
	GSC					–.148
	EI					.241
	ESC					–.162
LSC with	GI					.310
	GSC					.599
	EI					.188
	ESC					.711
	RI					–.198
RSC with	GI					.245
	GSC					.580
	EI					.139
	ESC					.683
	RI					–.136
	LSC					.831
RWI with	GI					.061
	GSC					–.065
	EI					.196
	ESC					–.113
	RI					.458
	LSC					–.130
	RSC					–.066
LI with	GI					.505
	GSC					.241
	EI					.265

(continued)

Table 8. Continued

	GI14 with GI2 EI10 with EI4	.539	.026
	ESC		.171
	RI		.182
	LSC		.319
	RSC		.264
	RWI		.180

Fit statistics	
$\chi^2(df, p)$	1,960.499 (628, <.001)
CFI	.900
SRMR	.052
RMSEA [90% CI]	.048 [.045, .050]

Note. CFA = confirmatory factor analysis; SE = standard error; RV = residual variance; V = variance; GI = Gender Identification; GSC = Gender Stigma Consciousness; EI = Ethnicity Identification; ESC = Ethnicity Stigma Consciousness; RI = Rural Identification; RSC = Rural Stigma Consciousness; LI = Low Income Identification; LSC = Low Income Stigma Consciousness; RWI = Reading And Writing Identification; CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval.
Item stems for girls of this scale can be found in Supplemental Appendix E. The final items for the stereotype threat scale for boys can be found in Supplemental Appendix D.

Table 9. Reliability Estimates (Internal Consistency) of Factors in the Stereotype Threat Scale.

Factors	Reliability estimates (Cronbach's α)
Gender Identification	.67
Gender Stigma Consciousness	.81
Ethnicity Identification	.77
Ethnicity Stigma Consciousness	.84
Rural Identification	.50
Rural Stigma Consciousness	.76
Low Income Identification	.69
Low Income Stigma Consciousness	.85
Reading and Writing Identification	.85

vulnerable with regard to growing up in an economically distressed area and how the rural context might influence student engagement, self-efficacy, growth mindset, and self-concept related to stereotype threat. We hope this discussion of our practices might inform other educational researchers to consider the multifaceted approach to designing and validating instruments focused on measuring these noncognitive constructs.

The measurement of noncognitive variables in the population of young gifted students requires considerable attention to the assessment of the reliability and validity of the instruments to be used. As the descriptions of the development and assessment of measures of self-efficacy, engagement, mindset, and stereotype threat reveal, the underlying factors may or may not parallel those identified in the general population or in older students. For example, the measure of self-efficacy evolved into an instrument measuring a single dimension, suggesting this construct may not have evolved into more specific underlying factors in younger children, and similarly, assessment of student engagement resulted in only two underlying factors (Behavioral Management and Emotional Engagement) rather than the four characterizing the measure in older children. Reliabilities of these factors, nonetheless, were well within acceptable ranges for use in analysis of group data.

However, we did find the best estimates of the underlying two-factor structure of the growth mindset scales did parallel those of the older population and the general population with very good reliabilities for affective measures. Also, Gender Identification, Gender Stigma Consciousness, Ethnicity Identification, and Ethnicity Stigma Consciousness of the scales for older students from the general population were again validated as were the added scales of Rural Identification, Rural Stigma Consciousness, Low Income Identification, Low Income Stigma Consciousness, and Reading and Writing Identification that were created for the particular gifted subpopulation under study. The estimates of the reliabilities of these subscales are also sufficiently high to warrant use in analysis of group data.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the U.S. Department of Education, under the Jacob K. Javits Gifted and Talented Students Education program, through Grant S206A140034-17.

ORCID iDs

Carolyn M. Callahan https://orcid.org/0000-0001-5056-1357
Christina M. Amspaugh https://orcid.org/0000-0002-0575-2098

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In the presentation of factor analysis of the data in the pilot tests, we discuss testing several models but present tables only for the model deemed best. The tables for all the other models tested are available from the authors upon request.
2. The enrichment program is specifically for gifted students. The sample is drawn from students selected on the basis of cognitive ability and creativity in academic areas as determined from an application process that includes student products and teacher rating of student cognitive ability. All applications are rated by trained doctoral students using specific rubrics to identify talent among the applicants.

References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*, 545–561. https://doi.org/10.1037/0022-0663.94.3.545

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385–390. https://doi.org/10.1111/1467-9280.00371

Andrade, H., Wang, X., Du, Y., & Akawi, R. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287–302.

Appleton, J. J., Christenson, S. L., Dongjin, K., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, *44*, 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125. https://doi.org/10.1006/jesp.2001.1491

Aronson, J., & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African-American students. *Psychological Science*, *15*, 829–836. https://doi.org/10.1111/j.0956-7976.2004.00763.x

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46. https://doi.org/10.1006/jesp.1998.1371

Aronson, J., & Steele, C. M. (2005). Stereotypes and the fragility of human competence, motivation, and self-concept. In C. Dweck & E. Elliot (Eds.), *Handbook of competence and motivation* (pp. 436–456). Guilford Press.

Aronson, J., Steele, C. M., Salinas, M. F., & Lustina, M. J. (1998). The effects of stereotype threat on the standardized test performance of college students. In E. Aronson (Ed.), *Readings about the social animal* (8th ed., pp. 415–430). Freeman.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5., pp. 307–337). Greenwich, CT: Information Age Publishing.

Bandura, A., Barbarnelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*, 1206–1222. https://doi.org/10.2307/1131888

Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing*, *12*(5), 461–469. https://doi.org/10.1177/1744987107079616

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bian, L., Leslie, S., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*, 389–391. https://doi.org/10.1126/science.aah6524

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Brown, R. P., & Josephs, R. A. (1999). *The importance of importance: The Mathematics Identification Questionnaire*. Unpublished manuscript, University of Texas at Austin.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*(6), 626–633. https://doi.org/10.1016/S0022-1031(03)00039-8

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Calkins, L., & Ehrenworth, M. (2016). Growing extraordinary writers: Leadership decisions to raise the level of writing across a school and a district. *The Reading Teacher*, *70*(1), 7–18. https://doi.org/10.1002/trtr.1499

Callahan, C. M. (2012). In closing. In R. F. Subotnik, A. Robinson, C. M. Callahan, E. J. Gubbins (Eds.), *Malleable minds: Translating insights from psychology and neuroscience to gifted education* (pp. 267–269). University of Connecticut, The National Research Center on the Gifted and Talented.

Chan, D. (2007). Positive and negative perfectionism among Chinese gifted students in Hong Kong: Their relationship to general self-efficacy and subjective well-being. *Journal for the Education of the Gifted*, *31*(1), 77–102.

Christenson, S. L., & Anderson, A. R. (2002). Commentary: The centrality of the learning context for students' academic enabler skills. *School Psychology Review*, *31*, 378–393.

Christenson, S. L., Reschly, A. L., Appleton, J. J., Berman, S., Spanjers, D., & Varro, P. (2008). Best practices in fostering student engagement. In A. Thomas & J. Grimes

(Eds.), *Best practices in school psychology* (5th ed., pp. 1099–1119). National Association of School Psychologists.

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer.

Clinkenbeard, P. R. (2012). Motivation and gifted students: Implications of theory and research. *Psychology in the Schools*, *49*, 622–630. https://doi.org/10.1002/pits.21628

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy and relatedness: A motivational analysis of self-system processes. In M. Gunnar & L. A. Sroufe (Eds.), *Minnesota Symposium on Child Psychology: Vol. 23. Self processes in development* (pp. 43–77). Lawrence Erlbaum.

Dai, D. Y., & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Child Quarterly*, *57*(3), 151–168. https://doi.org/10.1177/0016986213490020

Désert, M., Preaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education*, *24*, 207–218. https://doi.org/10.1007/BF03173012

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, *6*, 267–285.

Esparza, J., Schumow, L., & Schmidt, J. A. (2014). Growth mindset of seventh grade students in science. *NCSSSMST Journal*, *19*(1), 6–13.

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *High school reform and student management* (pp. 97–131). Springer. https://doi.org/10.1007/978-1-4614-2018-7_5

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. *2*, pp. 357–411). McGraw Hill.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59–109. https://doi.org/10.3102/00346543074001059

Gallagher, K. (2011). *Write like this: Teaching real-world writing through modeling and mentor texts*. Stenhouse.

Greenwald, A., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–28. https://doi.org/10.1037/0033-295X.102.1.4

Haimovitz, K., & Dweck, C. S. (2016). Parents' views of failure predict children's fixed and growth intelligence mind-sets. *Psychological Science*, *27*(6), 859–869. https://doi.org/10.1177/0956797616639727

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Pearson.

Hébert, T. P., & Beardsley, T. M. (2001). Jermaine: A critical case study of a gifted black child living in rural poverty. *Gifted Child Quarterly*, *45*, 85–102. https://doi.org/10.1177/001698620104500203

Hong, Y. Y., Chiu, C. Y., Dweck, C. S., Lin, D., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology*, *77*, 588–599. https://doi.org/10.1037/0022-3514.77.3.588

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, *17*, 63–84.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/1070551990954011

Jimerson, S. J., Campos, E., & Greif, J. L. (2003). Towards an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, *8*, 7–27. https://doi.org/10.1007/BF03340893

Jöreskog, K. G. (1999). How large can a standardized coefficient be? [Unpublished report]. http://www.statmodel.com/download/Joreskog.pdf

Landis, R. N., & Reschly, A. L. (2013). Reexamining gifted underachievement and dropout through the lens of student engagement. *Journal for the Education of the Gifted*, *36*, 220–249. https://doi.org/10.1177/0162353213480864

Lüftenegger, M., Kollmayer, M., Bergsmann, E., Jöstl, G., Spiel, C., & Schober, B. (2015). Mathematically gifted students and high achievement: The role of motivation and classroom structure. *High Ability Studies*, *26*, 227–243. https://doi.org/10.1080/13598139.2015.1095075

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318. https:// doi/10.1177/0146167292183006

Makel, M. C., Snyder, K. E., Thomas, C., Malone, P. S., & Pullatz, M. (2015). Gifted students' implicit beliefs about intelligence and giftedness. *Gifted Child Quarterly*, *59*, 203–212. https://doi.org/10.1177/0016986215599057

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, *27*, 486–493. https://doi.org/10.1016/j.appdev.2006.06.003

Melike, J. (2018). Investigation of gifted students' epistemological beliefs, self-efficacy beliefs and use of metacognition. *Journal for the Education of Gifted Young Scientists*, *6*(3), 1–10.

Mofield, E. L., & Peters, M. P. (2018). Mindset misconception? Comparing mindsets, perfectionism, and attitudes of achievement in gifted, advanced, and typical Students. *Gifted Child Quarterly*, *62*, 327–329.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*, 33–52. https://doi.org/10.1037/0022-3514.75.1.33

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards (English Language Arts)*.

Nguyen, D., & Ryan, A. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *The Journal of Applied Psychology*, *93*, 1314–1334. https://doi.org/10.1037/a0012702

Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology*, *21*, 325–344. https://doi.org/10.1006/ceps.1996.0025

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, *19*, 139–158.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*, 124–139. https://doi.org/10.1006/ceps.1998.0991

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English*, *42*(1), 104–120. https://doi.org/10.1037/a0024048

Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Journal of Educational Psychology*, *91*, 50–61. https://doi.org/10.1037/0022-0663.91.1.50

Park, S., Callahan, C. M., & Ryoo, J. H. (2016). Assessing gifted students' beliefs about intelligence with a psychometrically defensible scale. *Journal for the Education of the Gifted*, *39*, 288–314.

Picho, K., & Brown, S. W. (2011). Can stereotype threat be measured? A validation of the Social Identities and Attitudes Scale (SIAS). *Journal of Advanced Academics*, *22*, 374–411. https://doi.org/10.1177/1932202X1102200302

Picho, K., Rodriquez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, *153*, 299–233.

Reeve, J., & Lee, W. (2014). Students' classroom engagement produces longitudinal changes in classroom motivation. *Journal of Educational Psychology*, *106*, 527–540.

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: The case for the inclusion of student engagement variables. *Remedial and Special Education*, *27*, 276–292. https://doi.org/10.1177/07419325060270050301

Romeo, C., Master, A., Paunesku, D., Dweck, C. S., & Gross, J. J. (2014). Academic and emotional functioning in middle school: The role of implicit theories. *Emotion*, *14*, 227–234. https://doi.org/10.1037/a0035490

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology*, *87*, 386–398. https://doi.org/10.1037/0022-0663.87.3.386

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*, 91–100. https://doi.org/10.1037/0022-0663.81.1.91

Skinner, E. A. (1991). Development and perceived control: A dynamic model of action in context. In M. Gunnar & L. A. Sroufe (Eds.), *Minnesota Symposium on Child Development: Vol. 23. Self processes in development* (pp. 167–216). Lawrence Erlbaum.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*, 493–525. https://doi.org/10.1177/0013164408323233

Skinner, E. A., Pitzer, J. R., & Steele, J. S. (2016). Can student engagement serve as a motivational resource for academic coping, persistence, and learning during late elementary and early middle school? *Developmental Psychology*, *52*, 2099–2117.

Spencer, B., & Castano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research*, *20*, 418–432.

Spencer, S., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Steele, C. M., & Aronson, J. (1998). How stereotypes influence the standardized test performance of talented African American students. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 401–427). Brookings Institution.

Stewart, T. T. (2011). Demystifying what it means to be a writer: Making membership in the writing club possible. *Florida English Journal*, *45*(2), 5–10.

Strayhorn, T. L. (2009). The burden of proof: A quantitative study of high-achieving Black collegians. *Journal of African American Studies*, *13*(4), 375–387.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*, 3–54. https://doi.org/10.1177/1529100611418056

Upadyaya, K., & Salmela-Aro, K. (2013). Engagement with studies and work: Trajectories from postcomprehensive school education to higher education and work. *Emerging Adulthood*, *1*, 247–257. https://doi.org/10.1177/2167696813484299

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*, 1132–1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

Wasserberg, M. J. (2014). Stereotype threat effects on African American children in an urban elementary school. *Journal of Experimental Education*, *82*, 502–517.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R., & Schiefele, U. (2015). Development of achievement motivation and engagement. In *Handbook of child psychology and developmental science* (pp. 1–44). Wiley. https://doi.org/10.1002/9781118963418.childpsy320

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*, 51–59.

About the Authors

Carolyn M. Callahan, commonwealth professor of education at the University of Virginia, is currently the Co-Principal Investigator (PI) for the National Center for

Research in Gifted Education and PI for the Javits grant, Promoting PLACE. Her research interests include program evaluation, assessment, and underserved gifted students.

Amy Price Azano is an associate professor of education at Virginia Tech. Her research focuses on rural gifted education, adolescent literacy, and critical pedagogies of place. She is the Co-PI on a 5-year, $1.9 million grant, Promoting PLACE in Rural Schools, which serves to promote and support gifted education in high-poverty, rural school districts.

Sunhee Park, MA, PhD, is a data analyst in the Promoting PLACE project at the University of Virginia and a research consultant in the Measure of Cognitive Ability project at the University of Southern California. Her main areas of research include achievement motivation of gifted students and their talent-related career development, the development of creative problem-solving ability of gifted students, and assessment of affective characteristics and creativity.

Annalissa V. Brodersen is an education consultant and former research associate with the National Center for Research on Gifted Education and Project PLACE at the University of Virginia. Her research expertise is in gifted education policies and practices, experiences of teachers and students in high-poverty, rural schools, and the intersections between p-12 gifted education and higher education.

Melanie Caughey, PhD, is the visiting assistant professor for gifted and talented education at Cleveland State University. Her research interests include advanced secondary coursework and fidelity of implementation of curriculum.

Erika L. Bass is an assistant professor of English education at the University of Northern Iowa. Her research interests include writing instruction in secondary schools, engaging students in meaningful writing, and preparing future English teachers to empower their students through meaningful, deep instruction.

Christina M. Amspaugh is assistant professor and gifted education focus area lead in the Curry School of Education & Human Development at the University of Virginia. Her research interests include underrepresented gifted populations and the role of technology, public engagement, and open science in bridging the research–practice gap in gifted education.