Article

Journal of Psychoeducational Assessment
2019, Vol. 37(3) 307–319

DOI: 10.1177/0734282917743335
journals.sagepub.com/home/jpa
SAGE

Examining the Psychometric Properties of Maximally Efficient Items From the Social Skills Improvement System–Teacher Rating Scale

Christopher James Anthony[1] and James Clyde DiPerna[2]

Abstract

The current study examined the psychometric properties of a set of maximally efficient items (SMIs) from the Social Skills Improvement System–Teacher Rating Scale (SSIS-TRS). Sixty-three teachers rated 302 second through sixth graders on these SMIs, and several concurrent validity measures were administered. Results provided initial support for the validity of scores from the SMIs. Implications for practice and psychological research are discussed.

Keywords

social competence, social skills, problem behaviors, item response theory, Social Skills Improvement System rating scales

A growing body of research indicates that social competence is an important component of children's personal and academic success in school (e.g., Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Malecki & Elliot, 2002). Although social competence is a complex construct (Gresham, 2016; Whitcomb & Merrell, 2013), it is frequently conceptualized as including two broad categories of behavior: social skills and problem behaviors. Social skills refer to behaviors that make positive social interactions more likely and negative social interactions less likely. Problem behaviors in a social context refer to behaviors that inhibit the acquisition or application of social skills (Gresham & Elliott, 2008). A student's pattern of applying social skills and refraining from problem behaviors give rise to subjective judgments about their social competence (Gresham, 2016).

Research has linked social competence to several important student outcomes. For example, both social skills and problem behaviors have demonstrated positive and negative relationships, respectively, with academic outcomes including grades (Caprara et al., 2000; Wentzel, 1993) and scores on standardized measures of achievement (Malecki & Elliot, 2002; Turney & McLanahan, 2015; Wentzel, 1993). Furthermore, social competence is related to other important student

[1]Oklahoma State University, Stillwater, OK, USA
[2]Pennsylvania State University, State College, PA, USA

Corresponding Author:
Christopher James Anthony, School of Teaching, Learning, and Educational Sciences, Oklahoma State University, 223 Willard Hall, Stillwater, OK 74078-1001, USA.
Email: christopher.anthony@okstate.edu

outcomes such as peer rejection (Gresham, 1981), psychopathology (Cowen, Pederson, Babigian, Isso, & Trost, 1973; Segrin, 2000), school dropout, and chronic unemployment (Whitcomb & Merrell, 2013). Finally, social competence has been identified as an important outcome of the schooling process in its own right (DiPerna, Bailey, & Anthony, 2014). In recognition of its importance, researchers have developed a variety of intervention and prevention programs designed to promote the development of social competence (e.g., Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Greenberg et al., 2003).

Measurement of Social Competence

Although various methods exist for measuring social constructs (e.g., direct observation, interviews), behavior rating scales have been the most common method to assess social competence in research and clinical practice (Gresham, 2016). Strengths and weaknesses of measures are most appropriately evaluated relative to proposed uses and interpretation of scores (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). When assessing constructs such as social competence, clinicians and researchers must often attempt to balance strength of psychometric evidence with practical considerations such as measurement efficiency.

Measurement efficiency refers to optimizing the balance between the resources (time, costs) needed to complete a measure and the strength of the psychometric properties of its scores. Although longer measures tend to provide greater construct coverage and higher reliability estimates than shorter ones, these psychometric advantages may not justify the extra costs depending on the purpose of the assessment. Conversely, although brief measures can provide a cost-savings, they may yield scores with attenuated reliability (Widaman, Little, Preacher, & Sawalani, 2011) and limited construct validity.

In their review of social and emotional measures for young children (children ages birth to 5), Halle and Darling-Churchill (2016) identified six measures that exhibited psychometric properties adequate for broad usage in psychological research. Of those identified, five measured social competence, and the total number of items across these measures ranged from 57 (the Social Skills Rating Scale [SSRS]; Gresham & Elliot, 1990) to 166 (the Infant–Toddler Social and Emotional Assessment; Carter & Briggs-Gowan, 2000; median = 100 items). Although the length of these measures is related to their intended primary use for individual clinical decision-making, the number of items may be excessive for other applications in which brevity and efficiency are more important (e.g., large-scale screening, progress monitoring).

Given the dearth of brief social competence rating scales, many researchers have utilized ad hoc procedures for developing abbreviated versions of longer measures. For example, in a review of child care quality research, Zaslow et al. (2006) found that 27% of studies conducted from 1979 to 2005 modified socioemotional outcome measures but did not report any psychometric evidence for the modified version. Unfortunately, such practices reflect two unjustified assumptions: (a) psychometric properties of scores from validated measures automatically transfer to brief versions of those measures and (b) brief versions of longer measures require less evidence for reliability and validity (Smith, McCarthy, & Anderson, 2000). To address these problematic assumptions, Smith et al. (2000) called for full and rigorous validation of brief versions of existing measures conducted with independent samples.

Within the domain of social competence, one of the most prominent behavior rating scales is the Social Skills Improvement System–Teacher Rating Scale (SSIS-TRS; Gresham & Elliott, 2008). The SSIS-TRS is an updated version of the SSRS (Gresham & Elliot, 1990), and both measures have been widely used in research and practice (Gresham, Elliott, Vance, & Cook, 2011). From 2003 to 2008, for example, the SSRS was used in more than 50 peer-reviewed journal articles and 53 doctoral dissertations (Gresham et al., 2011). Despite its prominence, the

typical time needed to complete the SSIS-TRS (roughly 15-20 min; Gresham & Elliott, 2008) may prove challenging for some uses, such as targeted screening and progress monitoring.

To address this need, Anthony, DiPerna, and Lei (2016) utilized advanced methodology to identify a set of maximally efficient items (SMI) for each SSIS-TRS subscale. Specifically, they used item response theory (IRT; Hambleton, Swaminathan, & Rogers, 1991), an approach to scale development and refinement with particular advantages for improving measurement efficiency (e.g., Smith et al., 2000). Within an IRT framework, the probability of an item being rated within a certain response category (e.g., a rating of 2 on a Likert-type scale ranging from 1 to 4) is modeled as a function of the underlying trait being measured (e.g., cooperation) and various item-level parameters. Common item parameters estimated in polytomous IRT studies include threshold parameters, which determine the location of Option Characteristic Curves (OCCs) and discrimination parameters, which determine the slope of these OCCs.[1] Because these procedures are conducted at the item level rather than the test level, IRT facilitates identifying efficient items and streamlining existing measures (Edelen & Reeve, 2007). Specifically, IRT allows scale developers to identify at which levels (e.g., below average) of the target, trait (e.g., social competence) items are most reliable. Based on this information, scale developers can eliminate items that are not functioning efficiently for particular measurement purposes.

Consistent with this approach, Anthony et al. (2016) evaluated item information functions (IIFs; visual representations of the measurement precision across levels of the target trait) to select SSIS-TRS items that increased measurement precision at lower levels of the measured trait to target students at risk for difficulties in SSIS-TRS domains. They also balanced item selection with considerations of domain coverage, as well as other statistical criteria (e.g., ensuring that IRT assumptions were met for the resulting SMIs). Initial evidence for reliability and validity of SSIS-TRS SMI scores was promising; however, these findings were based on a single sample of teachers who completed the full-length SSIS-TRS on a single occasion. Although this study represented an important first step in potentially improving measurement efficiency of the SSIS-TRS for certain uses, further research with independent administration of SMIs is needed to further validate SMI scores (Smith et al., 2000).

Purpose and Hypotheses

Given the need for efficient measurement of social competence (Dirks, Treat, & Weersing, 2007; Whitcomb & Merrell, 2013), the purpose of this study was to examine the validity of the scores from the SSIS-TRS SMIs identified by Anthony et al. (2016). Consistent with this goal, several convergent and discriminant hypotheses were tested in the current study.

First, we generated hypotheses involving social skills and problem behaviors as measured by the SMIs. Because the SMIs and full-length SSIS-TRS intend to measure the same constructs, we hypothesized that their scores would be strongly positively related. In addition, we specified hypotheses about the interrelationships between SMI constructs. Specifically, because they measure related social skills and problem behaviors constructs (Gresham & Elliott, 2008), we hypothesized that social skills as measured by the SMIs would be moderately to strongly positively interrelated. Similarly, we hypothesized that problem behaviors as measured by the SMIs would be moderately to strongly positively interrelated. Finally, we hypothesized that social skills and problem behaviors as measured by the SMIs would be moderately to strongly negatively related.

Second, given previous research examining the relationship between social behavior and academic enablers (e.g., motivation and engagement; Demaray & Jenkins, 2011; DiPerna & Elliott, 2000; DiPerna, Volpe, & Elliott, 2002, 2005; McDermott, Leigh, & Perry, 2002), we predicted that social skills and problem behaviors would be moderately related to academic enablers. We predicted that these relationships would be positive for social skills and negative for problem behaviors.

Table 1. Demographic Characteristics, Academic Skills, Motivation, and Engagement of Participating Sample (N = 302).

	Participants	U.S. student population[a]
Female	49	49
Race		
White	77	50
Black	12	16
Hispanic	6	25
Other	4	9
Educational status		
General education	90	86
Special education	9	14
STAR		
Reading	520.15 (263.44)	—
Mathematics	639.54 (54)	—
ACES		
Motivation	3.57 (0.98)	—
Engagement	3.84 (0.90)	—

Note. Demographic characteristics reported as percentages. For some variables, percentages do not add to 100 due to missing data and/or rounding. STAR and ACES data reported as means (standard deviations). STAR Reading and Mathematics n = 165 and 162 students, respectively. ACES = Academic Competence Evaluation Scales; STAR = Standardized Tests for the Assessment of Reading.

[a]Based on the 2013-2014 Civil Rights Data Collection survey of public schools (U.S. Department of Education, Office for Civil Rights, 2016).

Finally, because construct inferences are informed not only by moderate/strong relationships with similar constructs but also by weaker relationships with less closely related constructs (AERA, APA, & NCME, 2014; Cronbach & Meehl, 1955), we generated discriminant validity hypotheses about the relationships between constructs on the SMIs and directly measured academic skills (reading and mathematics). Based on previous associations between social skills, problem behaviors, and academic achievement (e.g., Caemmerer & Keith, 2015; DiPerna et al., 2002, 2005; Fleming et al., 2005; Malecki & Elliot, 2002; Oberle, Schonert-Reichl, Hertzman, & Zumbo, 2014), we predicted that social skills and problem behaviors would demonstrate small relationships with directly measured academic achievement. Similar to our hypotheses for academic motivation and engagement, we predicted that these small relationships would be positive for social skills and negative for problem behaviors.

Method

Participants

The sample for this study consisted of 302 students in Grades 2 to 6 across seven public schools in the Mid-Atlantic region of the United States. Participants ranged in age from 6.67 to 12.33 years (median = 8.93 years). With regard to grade, 22% of students were in the second grade, 26% in third, 23% in fourth, 16% in fifth, and 13% in sixth. As shown in Table 1, participants were evenly distributed across gender, and a majority (77%) were White. Approximately 9% of the student participants were receiving special education services. Participating classrooms (n = 63) had a median enrollment of 23 students, were approximately evenly distributed across schools, and had a median of four (range = 1-11) students participate in the project. Teachers (n = 63) were predominantly White (98%) and female (85%). All teachers spoke English as their

Table 2. Number of Items, Mean Item Ratings, and Correlations Between SMI and SSIS-TRS Scores.

	SMI			SSIS-TRS			
Scale/subscale	*n*	*M*	*SD*	*n*	*M*	*SD*	*r* (95% CI)
Social skills							
Communication	3	2.46	0.57	7	2.45	0.58	.76 [.70, .83]
Cooperation	4	2.3	0.65	6	2.25	0.65	.80 [.75, .86]
Assertion	4	2.28	0.64	7	2.08	0.59	.65 [.58, .73]
Responsibility	4	2.43	0.61	6	2.38	0.63	.80 [.75, .85]
Empathy	4	2.36	0.69	6	2.31	0.66	.79 [.74, .84]
Engagement	4	2.41	0.59	7	2.36	0.61	.81 [.75, .86]
Self-control	4	2.38	0.69	7	2.26	0.67	.78 [.73, .83]
Total	27	2.37	0.55	46	2.30	0.54	.85 [.81, .89]
Problem behaviors							
Externalizing	7	0.39	0.51	12	0.40	0.53	.82 [.76, .88]
Bullying	4	0.25	0.49	5	0.27	0.50	.73 [.60, .86]
Hyperactivity/ Inattention	5	0.51	0.56	7	0.58	0.63	.79 [.72, .86]
Internalizing	4	0.40	0.56	7	0.43	0.52	.75 [.68, .83]
Total	14	0.40	0.46	25	0.43	0.48	.80 [.73, .87]

Note. The sum of Problem Behaviors subscale items is greater than the total number of Problem Behavior items because several are included on multiple subscales. All items were rated on a scale from 1 to 4. SMI = Set of Maximally Efficient Items; SSIS-TRS = Social Skills Improvement System–Teacher Rating Scale; CI = confidence interval.

primary language, and 79% had received a bachelor's degree as their highest degree. Teachers had taught from 6 to 37 years (median = 15.5 years) in total and from 2 to 21 years at their current grade level (median = 8 years).

Measures

Several measures were included to evaluate the psychometric adequacy of scores from the SMIs. Specifically, following recommendations of Smith et al. (2000), the full SSIS-TRS was completed separately from the SMIs to examine construct validity of SMI ratings. Other concurrent validity measures assessed participants' academic enablers (Academic Competence Evaluation Scales [ACES]; DiPerna & Elliott, 2000) and academic achievement (STAR Reading and Mathematics; Renaissance Learning, 2007). Descriptive statistics for validity measures and SMI/SSIS-TRS scores can be found in Tables 1 and 2, respectively.

SSIS-TRS. The SSIS-TRS (Gresham & Elliott, 2008) is a behavior rating scale measuring student-level social skills and problem behaviors. The SSIS-TRS produces seven subscale scores for social skills (Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control), four for problem behaviors (Externalizing, Bullying, Hyperactivity/Inattention, and Internalizing), and two broad scale scores (Social Skills and Problem Behaviors). In addition, the SSIS-TRS includes Autism Spectrum and Academic Competence subscales that were not used in this study. The SSIS-TRS consists of 76 items (excluding the Autism Spectrum and Academic Competence subscales) rated on a 4-point Likert-type scale ranging from 0 (*never*) to 3 (*almost always*). With regard to evidence for reliability of scores, the SSIS-TRS manual reports internal consistency estimates > .90 for SSIS-TRS scale scores and > .80 for subscale

scores. With regard to evidence for validity, scores from the SSIS-TRS have been reported to correlate as expected with scores from various validity measures (e.g., the Behavior Assessment System for Children–Second Edition; Reynolds & Kamphaus, 2004) both in the SSIS-TRS manual (Gresham & Elliott, 2008) and in external studies (e.g., Gresham, Elliott, Cook, Vance, & Kettler, 2010; Gresham et al., 2011).

SSIS-TRS SMIs. The SMIs identified by Anthony et al. (2016) include 41 (58%) of the items on the Social Skills and Problem Behaviors scales from the full-length SSIS-TRS. Anthony et al. (2016) found the SMIs to exhibit adequate internal consistency with Cronbach's alpha ranging from .84 to .97 (median = .90) and stability coefficients ranging from .57 to .72 (median = .66; average interval length = 4.5 months) across scales and subscales. Furthermore, SMIs related as expected with other scores from academic and behavioral measures. Specifically, Anthony et al. (2016) reported validity coefficients for Social Skills scale and subscale SMIs ranged from .11 to .36 (median = .27) for STAR Reading and Mathematics scores, and from .46 to .74 (median = .63) for ACES Motivation and Engagement scores. Furthermore, validity coefficients for Problem Behaviors scale and subscale SMI scores ranged from −.31 to −.14 (median = −.24) for STAR Reading and Mathematics scores, and from −.60 to −.19 (median = −.43) for ACES Motivation and Engagement scores. Results also indicated small decreases in the magnitude of reliability and validity estimates when compared with those calculated with full-length SSIS-TRS scores. Anthony et al. (2016), however, examined the psychometric properties of SMI data drawn from administrations of the full-length SSIS-TRS. For the current study, the SMIs were administered separately from the full-length SSIS-TRS to estimate the psychometric properties of SMI-based scores. The independent administration of the SSIS-TRS and SMIs was chosen to ensure that comparisons between validity coefficients calculated with the SMIs and SSIS-TRS would not be complicated by sharing items rated by the same individuals, at the same time. This approach was recommended by Smith et al. (2000) for such validation projects.

Academic Competence Evaluation Scales–Teacher Form (ACES-TF). The ACES-TF (DiPerna & Elliott, 2000) Motivation and Engagement subscales were used as concurrent validity measures in this study. Items on these subscales are rated on a 5-point Likert-type scale ranging from 1 (*never*) to 5 (*almost always*). As reported in the test manual, Cronbach's alpha levels for scores from these subscales ranged from .94 to .95 for the Engagement subscale and from .98 to .97 for the Motivation subscale across subsamples. Test–retest reliability coefficients for the full sample were .92 and .96 for the Engagement and Motivation subscales, respectively. With regard to validity evidence, scores from these subscales have demonstrated expected relationships with similar constructs (DiPerna & Elliott, 2000). For example, DiPerna and Elliott (2000) reported large positive correlations and medium to large negative correlations between ACES Motivation and Engagements scores and social skills and problem behaviors, respectively.

STAR Reading and Mathematics. The STAR Reading and Mathematics assessments (Renaissance Learning Inc., 2007) were also used as a validity measure in this study. These data were collected with a subsample of third-, fifth-, and sixth-grade participants (n = 165 for Reading, n = 162 for Mathematics). These computer adaptive tests measure student skills in reading and mathematics. For STAR reading scores, internal consistency estimates ranged from .89 to .91 (median = .90) and test–retest reliability coefficients ranged from .82 to .89 (median = .83; U.S. Department of Education, National Center on Response to Intervention, 2010). Corresponding coefficients ranged from .79 to .83 (internal consistency; median = .81) and from .73 to .79 (test–retest reliability; median = .74) for STAR Mathematics scores. With regard to validity evidence, STAR Reading and Mathematics scores have been moderately to strongly positively related to other

direct measures of academic achievement in various studies (U.S. Department of Education, National Center on Response to Intervention, 2010).

Procedures

Data were collected during the latter half of 1 school year. From late February to early April, teachers rated students on the SSIS-TRS and the ACES Motivation and Engagement subscales, and participating students completed the STAR Reading and Mathematics. Teachers then completed the SMI items for each participating student from their classroom from May to June. The average interval between SSIS-TRS and SMI administration was approximately 10 weeks (range = 6-12 weeks). This order of administration was chosen to maximize the independence of SSIS and SMI ratings (Smith et al., 2000). Participating teachers received monetary compensation for completing the SSIS-TRS, ACES Motivation and Engagement subscales, and SMIs.

Data Analysis

Several data analyses were conducted to examine the psychometric characteristics of scores from the SMIs. All SMI scores were calculated by taking the mean of the item ratings for each scale/subscale. First, correlation coefficients were computed between SMI scores and SSIS-TRS scores. Next, scale and subscale SMI intercorrelations were computed. Finally, correlation coefficients were computed between SMI scores and the aforementioned validity measures (ACES Engagement and Motivation subscales; STAR Reading and Mathematics). In addition, correlations were calculated between the validity measures and the SSIS-TRS for comparative purposes. All correlation coefficients were adjusted for clustering within teacher/classroom using the method outlined by Lorenz, Datta, and Harkema (2011). This method is equivalent to within cluster resampling, a procedure by which correlations are iteratively computed on nonnested data (i.e., a single observation is selected from each cluster) and averaged. These procedures were conducted with R version 3.3.1 (2016) and 95% confidence intervals were calculated for each correlation. Validity coefficients calculated with SMI and SSIS-TRS scores were statistically compared using procedures outlined by Steiger (1980) with an a priori alpha of .05 (adjusted via a Bonferroni correction within validity measure).

Results

As shown in Table 2, correlations between the social skills SMIs and SSIS-TRS Social Skills scale and subscale scores ranged from .65 to .85 (median = .79), and correlations between problem behaviors SMIs and SSIS-TRS Problem Behavior scale and subscale scores ranged from .73 to .82 (median = .79). Social skills subscale SMI intercorrelations (Table 3) ranged from .47 to .86 (median = .76) and problem behaviors subscale SMI intercorrelations ranged from .46 to .92 (median = .65). Intercorrelations between social skills and problem behaviors subscale SMIs ranged from −.75 to −.21 (median = −.59).

Concurrent validity correlations between social skills SMIs and ACES scores ranged from .49 to .70 (median = .62) for Motivation and .43 to .63 (median = .56) for Engagement (Table 4). Correlations between problem behaviors SMI scores and ACES scores ranged from −.55 to −.35 (median = −.48) for Motivation and −.45 to −.19 (median = −.39) for Engagement (Table 4). Correlations between social skills SMI scores and STAR scores ranged from .06 to .23 (median = .16) for Reading and .11 to .31 (median = .25) for Mathematics (Table 5). Correlations between problem behaviors SMI scores and STAR scores ranged from −.17 to −.10 (median = −.16) for Reading and −.30 to −.19 (median = −.28) for Mathematics.

Table 3. Subscale SMI Intercorrelations.

Scale/subscale SMI	1	2	3	4	5	6	7	8	9	10	11
1. Communication	—										
2. Cooperation	.76	—									
3. Assertion	.65	.47	—								
4. Responsibility	.85	.80	.53	—							
5. Empathy	.77	.68	.62	.80	—						
6. Engagement	.86	.71	.72	.82	.75	—					
7. Self-control	.79	.72	.47	.86	.77	.77	—				
8. Externalizing	−.61	−.68	−.26	−.72	−.63	−.55	−.74	—			
9. Bullying	−.51	−.51	−.21	−.61	−.58	−.47	−.64	.91	—		
10. Hyperactivity/Inattention	−.61	−.75	−.30	−.72	−.60	−.56	−.70	.92	.73	—	
11. Internalizing	−.60	−.51	−.51	−.55	−.43	−.68	−.49	.51	.46	.56	—

Note. SMI = set of maximally efficient items.

As shown in Tables 4 and 5, correlations between SSIS-TRS scores and validity measures were similar to the correlations between the SMIs and the validity measures. Validity coefficients were slightly smaller in magnitude when calculated with SMI scores relative to SSIS-TRS scores (median difference in magnitude across all validity coefficients = .06). Although a small number of validity coefficient differences were statistically significant (see Table 4), the magnitudes of these differences were small, ranging from .09 to .12 (median = .11).

Discussion

Overall, results of the current study provide further validity evidence for scores based on the SSIS-TRS SMIs identified by Anthony et al. (2016). First, scores from the SMIs demonstrated hypothesized relationships with scores from validity measures. Correlations between SMI and SSIS-TRS scores were large; SMI intercorrelations were generally medium to large; correlations between SMI scores and ACES scores were generally medium to large; and correlations between SMI scores and STAR scores were generally small. In line with expectations, relationships between social skills SMI scores, motivation, engagement, and academic skills were positive while corresponding relationships involving problem behaviors SMI scores were negative and SMI score intercorrelations were in the expected direction. Furthermore, there were very slight differences in magnitude between validity coefficients when calculated with SMI and SSIS-TRS scores. These results indicate that the constructs represented by SMI scores and the SSIS-TRS scores demonstrate very similar relationships with the validity constructs included in this study. Each of these findings is consistent with those noted by Anthony et al. (2016).

In addition to small differences between validity coefficients, the observed pattern of magnitudes of SMI validity coefficients is generally consistent with expectations. Specifically, the average magnitude of correlations was highest between SMI and SSIS-TRS scores. The next highest average magnitude of correlations was between SMI and ACES scores. Finally, the lowest average magnitude of correlations was between SMI and STAR scores. This pattern is what would be expected given the influences of shared trait and method variance (e.g., Campbell & Fiske, 1959) and theoretical relationships between constructs. Specifically, those measures that shared trait and method (i.e., teacher rating scale) variance (SMI and SSIS-TRS; $.65 \leq |r| \leq .85$; median = .79) evidenced stronger correlations than those that shared method variance but not trait variance (SMI and ACES; $.19 \leq |r| \leq .75$; median = .56) which were in turn stronger than those that shared neither method nor trait variance (SMI and STAR; $.02 \leq |r| \leq .37$; median = .23).

Table 4. Convergent and Discriminant Relationships Between ACES Motivation and Engagement Scores With the SMIs and SSIS-TRS.

	ACES Motivation		ACES Engagement	
	SMI (95% CI)	SSIS-TRS (95% CI)	SMI (95% CI)	SSIS-TRS (95% CI)
Social skills				
Communication	.62 [.54, .71]	.66 [.58, .73]	.57 [.48, .65]	.63 [.54, .71]
Cooperation	.70 [.64, .76]	.75 [.69, .81]	.56 [.46, .66]	.57 [.48, .66]
Assertion	.49 [.38, .60]	.48 [.36, .60]	.62 [.54, .70]	.66 [.59, .73]
Responsibility	.63 [.56, .71]	.68 [.62, .75]	.48 [.37, .59]	.53 [.43, .63]
Empathy	.52* [.43, .61]	.61 [.53, .69]	.49* [.38, .60]	.58 [.48, .68]
Engagement	.61 [.53, .69]	.65 [.57, .74]	.63 [.55, .72]	.70 [.62, .78]
Self-control	.53 [.44, .62]	.57 [.49, .66]	.43 [.30, .55]	.45 [.33, .57]
Total	.67 [.60, .74]	.72 [.65, .79]	.62 [.53, .70]	.68 [.60, .76]
Problem behaviors				
Externalizing	−.48 [−.62, −.34]	−.56 [−.63, −.48]	−.29 [−.43, −.15]	−.36 [−.47, −.25]
Bullying	−.35 [−.49, −.20]	−.40 [−.51, −.29]	−.19 [−.33, −.05]	−.27 [−.38, −.16]
Hyperactivity/ Inattention	−.55* [−.67, −.44]	−.67 [−.73, −.60]	−.39* [−.52, −.25]	−.49 [−.59, −.39]
Internalizing	−.44* [−.57, −.32]	−.55 [−.64, −.46]	−.45* [−.58, −.32]	−.57 [−.67, −.47]
Total	−.54* [−.69, −.40]	−.65 [−.72, −.57]	−.41* [−.56, −.26]	−.51 [−.62, −.41]

Note. ACES = Academic Competence Evaluation Scales; SMI = set of maximally efficient items; CI = confidence interval; SSIS-TRS = Social Skills Improvement System–Teacher Rating Scale.
*$p < .05$ for difference between SMI and SSIS-TRS validity coefficient.

As such, the pattern of findings provides further evidence of the construct validity of scores from the SMIs. As with the other general findings in this study, this pattern of findings replicates the pattern observed by Anthony et al. (2016).

From a practical perspective, a measure based on the SMIs would include 58% of the social skills and problem behaviors items of the SSIS-TRS. The SSIS-TRS manual indicates that typical time for completion of a SSIS-TRS is 15 to 20 min. Based on this estimate, a measure based on the SMIs would save roughly 6 to 8 min per student. Such time-savings could quickly compound in contexts requiring large numbers of ratings (e.g., 300 ratings would save 30 to 40 hr of rater time). Given the positive psychometric characteristics evidenced in this study, such time-savings would not likely lead to substantial decrement of the quality of inferences based on SMI scores.

There would be several other strengths of an abbreviated measure based on the SMIs. First, the procedures used to identify the SMIs are more systematic and data-informed than the ad hoc procedures often used to develop brief versions of existing measures (Zaslow et al., 2006). Furthermore, the SMIs have undergone evaluation with an independent sample and administration, which is uncommon in the short form development literature (Smith et al., 2000). Thus, the SMIs represent a balance between brevity, content coverage, and psychometric adequacy. These advantages indicate the potential of the SMIs should they be fully developed into an abbreviated version of the SSIS-TRS.

There are several potential clinical applications for an abbreviated version of the SSIS-TRS based on the SMIs. Specifically, such a measure could offer clinicians a tool for situations in which brevity and efficiency of measurement are of greater importance than extensive content coverage. Although the current study did not evaluate score precision, Anthony et al. (2016) focused their development on an at-risk range (i.e., lower levels of the target trait) in

Table 5. Convergent and Discriminant Relationships Between STAR Reading and Mathematics Scores With the SMIs and SSIS-TRS.

	STAR Reading		STAR Mathematics	
	SMI (95% CI)	SSIS-TRS (95% CI)	SMI (95% CI)	SSIS-TRS (95% CI)
Social skills				
Communication	.22 [.06, .38]	.26 [.09, .43]	.31 [.16, .47]	.30 [.11, .49]
Cooperation	.23 [.09, .37]	.30 [.14, .46]	.31 [.16, .46]	.37 [.21, .54]
Assertion	.06 [−.11, .23]	−.02 [−.18, .14]	.11 [−.12, .33]	.10 [−.10, .30]
Responsibility	.18 [−.01, .37]	.24 [.07, .40]	.26 [.08, .45]	.30 [.14, .47]
Empathy	.07 [−.11, .26]	.09 [−.08, .26]	.16 [−.03, .36]	.15 [−.05, .35]
Engagement	.15 [.03, .28]	.19 [.02, .36]	.24 [.10, .38]	.22 [.02, .42]
Self-control	.12 [−.05, .28]	.17 [.00, .34]	.23 [.06, .40]	.27 [.10, .44]
Total	.16 [.00, .33]	.20 [.03, .37]	.26 [.09, .43]	.28 [.09, .47]
Problem behaviors				
Externalizing	−.17 [−.33, .00]	−.25 [−.38, −.12]	−.28 [−.45, −.12]	−.34 [−.50, −.18]
Bullying	−.10 [−.24, .04]	−.21 [−.32, −.11]	−.19 [−.34, −.04]	−.28 [−.43, −.12]
Hyperactivity/ Inattention	−.17 [−.37, .03]	−.27 [−.45, −.09]	−.29 [−.47, −.11]	−.34 [−.52, −.16]
Internalizing	−.11 [−.24, .03]	−.19 [−.35, −.04]	−.26 [−.41, −.10]	−.28 [−.44, −.12]
Total	−.16 [−.33, .01]	−.27 [−.42, −.13]	−.30 [−.47, −.14]	−.37 [−.53, −.21]

Note. No differences between SMI and SSIS-TRS validity coefficients were statistically significant. ACES = Academic Competence Evaluation Scales; SMI = set of maximally efficient items; SSIS-TRS = Social Skills Improvement System–Teacher Rating Scale; CI = confidence interval; STAR = Standardized Tests for the Assessment of Reading.

the identification of the SMIs. As such, a brief scale based on the SMIs would likely function optimally in clinical situations in which expected levels of social competence fall in this range. Such situations include targeted screening of at-risk populations and periodic monitoring of individuals or groups receiving interventions to ameliorate social skills or problem behavior difficulties. Such a measure might also function adequately as a second-level assessment (i.e., a follow-up measure administered after an initial universal screener) within systematic screening systems (e.g., the Systematic Screening for Behavior Disorders [SSBD]; Walker, Severson, & Feil, 2014).

An abbreviated version of the SSIS-TRS based on the SMIs also has potential applications in research. As mentioned previously, there has been increasing emphasis on the development and evaluation of social and emotional learning programs (e.g., Durlak et al., 2011; Greenberg et al., 2003). Given the need for efficient outcome measures in such studies, the SMIs in particular and the IRT methods used to develop them in general hold promise to promote further research and limit difficulties such as attrition, decreased statistical power, and questionable validity due to excessive participant burden (Galesic & Bosnjak, 2009).

Limitations and Directions for Future Research

There are several limitations of this study to address in future research. For example, the interval between collection of the validity measures and the SMIs data was approximately 10 weeks. This lag may have attenuated validity coefficients relative to those collected with shorter lags between data collection. Future research should examine coefficients based on data collected in closer time proximity (e.g., 1 week). Furthermore, the current sample was not representative of the broader national population with regard to several key demographic variables (e.g., race) and was

limited to students in Grades 2 to 6. As such, future research should examine SMI psychometric properties with a sample that closely mirrors the national K-12 student population.

Beyond addressing the limitations of the current study, there are additional lines of validity evidence to examine for the SMIs. Although Anthony et al. (2016) considered content validity when identifying the SMIs, expert ratings of content coverage would provide additional insight regarding SMI content validity. In addition, future research is necessary to test the structure of the SMIs to see if it parallels the theoretical framework that guided the development of the SSIS-TRS. Given the importance of accurate classification and prediction of outcomes by screening measures (Smith et al., 2000; Youngstrom, 2013), these lines of evidence also should be examined in future SMI studies. Finally, collection of teachers' perspectives regarding usability, feasibility, and applicability, would provide insight regarding the social validity of the SMIs.

Conclusion

The current study expands the evidence base for the SMIs identified by Anthony et al. (2016). Specifically, current evidence supports the construct validity of SMI scores. In addition to producing scores with evidence for validity, the SMIs are efficient. Scores from the SMIs appear to measure the same constructs as the SSIS-TRS with roughly 50% fewer items than the SSIS-TRS. These findings indicate that the SMIs examined in this study potentially could be used to inform future development of an abbreviated version of the SSIS-TRS for targeted screening, progress monitoring, or research applications. When considered in conjunction with the results reported by Anthony et al. (2016), this study further supports the use of IRT methodology to improve the efficient measurement of social competence and other constructs. Given the foundational importance of measurement to the scientific and clinical examination of socioemotional and behavioral constructs, such developments are necessary to support continued growth of psychological research and practice.

Declaration of Conflicting Interests

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

Funding

The author(s) received the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A090438 and R305B090007 to The Pennsylvania State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Note

1. For further information about item response theory (IRT) in general, readers are referred to de Ayala (2009) or Hambleton, Swaminathan, and Rogers (1991). For further information about polytomous IRT, readers are referred to Nering and Ostini (2010).

References

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

Anthony, C. J., DiPerna, J. C., & Lei, P.-W. (2016). Maximizing measurement efficiency of behavior rating scales using item response theory: An example with the Social Skills Improvement System–Teacher Rating Scale. *Journal of School Psychology*, *55*, 57-69. doi:10.1016/j.jsp.2015.12.005

Caemmerer, J. M., & Keith, T. Z. (2015). Longitudinal, reciprocal effects of social skills and achievement from kindergarten to eighth grade. *Journal of School Psychology*, *53*, 265-281.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Caprara, G., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. (2000). Prosocial foundations of children's academic achievement. *Psychological Science*, *11*, 302-306.

Carter, A. S., & Briggs-Gowan, M. J. (2000). *Manual of the Infant-Toddler Social-Emotional Assessment*. New Haven, CT: Yale University.

Cowen, E. L., Pederson, A., Babigian, H., Isso, L. D., & Trost, M. A. (1973). Long-term follow-up of early detected vulnerable children. *Journal of Consulting and Clinical Psychology*, *41*, 438-446.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Demaray, M. K., & Jenkins, L. N. (2011). Relations among academic enablers and academic achievement in children with and without high levels of parent-rated symptoms of inattention, impulsivity, and hyperactivity. *Psychology in the Schools*, *48*, 573-586. doi:10.1002/pits.20578

DiPerna, J. C., Bailey, C. G., & Anthony, C. J. (2014). Broadband screening of academic and social behavior. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Identification, implementation, and interpretation*. (pp. 223-248). Washington, DC: APA Books.

DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: The Psychological Corporation.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review*, *31*, 298-312.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2005). A model of academic enablers and mathematics achievement in the elementary grades. *Journal of School Psychology*, *43*, 379-392.

Dirks, M. A., Treat, T. A., & Weersing, V. R. (2007). Integrating theoretical, measurement, and intervention models of youth social competence. *Clinical Psychology Review*, *27*, 327-347. doi:10.1016/j.cpr.2006.11.002

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*, 405-432. doi:10.1111/j.1467-8624.2010.01564.x

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*, 5-18.

Fleming, C. B., Haggerty, K. P., Catalano, R. F., Harachi, T. W., Mazza, J. J., & Gruman, D. H. (2005). Do social and behavioral characteristics targeted by preventive interventions predict standardized test scores and grades? *Journal of School Health*, *75*, 342-349.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*, 349-360. doi:10.1093/poq/nfp031

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, *58*, 466-474.

Gresham, F. M. (1981). Social skills training with handicapped children: A review. *Review of Educational Research*, *51*, 139-176.

Gresham, F. M. (2016). Social skills assessment and intervention for children and youth. *Cambridge Journal of Education*, *46*, 319-332. doi:10.1080/0305764X.2016.1195788

Gresham, F. M., & Elliot, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System–Rating Scales*. Minneapolis, MN: Pearson Assessments.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System–Rating Scales. *Psychological Assessment*, *22*, 157-166. doi:10.1037/a0018124

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, *26*, 27-44. doi:10.1037/a0022662

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology*, *45*, 8-18. doi:10.1016/j.appdev.2016.02.003

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Lorenz, D. J., Datta, S., & Harkema, S. J. (2011). Marginal association measures for clustered data. *Statistics in Medicine*, *30*, 3181-3191.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*, 1-23. doi:10.1521/scpq.17.1.1.19902

McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the preschool learning behaviors scale. *Psychology in the Schools*, *39*, 353-365. doi:10.1002/pits.10036

Nering, M. L., & Ostini, R. (Eds.). (2010). *Handbook of polytomous item response theory models*. New York, NY: Routledge.

Oberle, E., Schonert-Reichl, K. A., Hertzman, C., & Zumbo, B. D. (2014). Social–emotional competencies make the grade: Predicting academic success in early adolescence. *Journal of Applied Developmental Psychology*, *35*, 138-147.

R Development Core Team. (2016). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Available from https://www.R-project.org/

Renaissance Learning, Inc. (2007). *Understanding STAR assessments*. Wisconsin Rapids, WI: Author.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Bloomington, MN: Pearson Assessments.

Segrin, C. (2000). Social skills deficits associated with depression. *Clinical Psychology Review*, *20*, 379-403.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102-111. doi:10.1037/1040-3590.12.1.102

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251.

Turney, K., & McLanahan, S. (2015). The academic consequences of early childhood problem behaviors. *Social Science Research*, *54*, 131-145. doi:10.1016/j.ssresearch.2015.06.022

U.S. Department of Education, National Center on Response to Intervention. (2010). *Review of progress monitoring tools* [Review of STAR Math]. Washington, DC: Author.

U.S. Department of Education, Office for Civil Rights. (2016). *2013-2014 Civil rights data collection: A first look*. Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf

Walker, H. M., Severson, H. H., & Feil, E. G. (2014). *Systematic screening for behavior disorders (SSBD) technical manual: Universal screening for preK–9* (2nd ed.). Eugene, OR: Pacific Northwest.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*, 357-364.

Whitcomb, S., & Merrell, K. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists* (pp. 39-61). Washington, DC: American Psychological Association.

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology*, *42*, 139-159. doi:10.1080/15374416.2012.736358

Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., & Geyelin, N. M. (2006). Child outcome measures in the study of child care quality. *Evaluation Review*, *30*, 577-610. doi:10.1177/0193841X06291529