

BRIEF REPORT

The SSIS SEL Brief Scales—Student Form: Initial Development and Validation

Christopher J. Anthony
University of FloridaStephen N. Elliott
Arizona State UniversityJames C. DiPerna and Pui-Wa Lei
The Pennsylvania State University

The SSIS SEL Brief Scales—Student Form (SSIS SEL*b*-S) was developed to create an efficient assessment of students' social and emotional learning (SEL). Using item response theory with ratings from 800 students in Grades 3–12 from the standardization sample, 20 items were selected from the full-length SSIS SEL Rating Form - Student to maximize score information and rating efficiency. After identifying items for the SSIS SEL*b*-S, we conducted several reliability and validity analyses. These analyses provided initial support for the use of the SSIS SEL*b*-S for low-stakes decision making contexts. As such, the SSIS SEL*b*-S holds promise for incorporating the perspectives of students ages 8–18 into assessments of their SEL competencies.

Impact and Implications





There is increasing need for efficient assessments targeting students' social and emotional learning (SEL) competencies to support school-based SEL programming. To address this need, the current study used advanced psychometric approaches to develop the SSIS SEL Brief Scales - Student Form, a brief version of the prominent SSIS SEL Rating Form—Student. The resulting measure holds promise to extend and support SEL programming and incorporate students' self-perceptions into the SEL assessment and intervention process.

Keywords: social and emotional learning, student self-ratings, item response theory

Increasing attention is being paid to supporting students' social and emotional learning (SEL) skills in schools across the United States and internationally. Among different types of SEL assessments, self-report assessments commonly are used and considered “irreplaceable sources of information about children's views of themselves” (McKown, 2017, p. 323). Indeed, for many SEL and related domains, student self-report is viewed as the most valid source of information because it encompasses emotions and skills to which only students themselves have access (e.g., internalizing

concerns; Smith, 2007). Other common SEL competencies to which students have unique insight include understanding and managing one's thoughts, emotions and behaviors as well as empathizing and relating with others (Weissberg, Durlak, Domitrovich, & Gullotta, 2015). Although there are many frameworks that enumerate these and other important SEL domains, one in particular—the framework developed by the Collaborative for Academic Social and Emotional Learning (CASEL, 2012; see Table 1)—has become prominent in research and practice as well as highly influential in SEL-focused policy (e.g., Eklund, Kilpatrick, Kilgus, & Haider, 2018).

Despite the growing prominence of SEL in general and the CASEL framework in particular, implementation challenges remain, especially in the area of self-report assessment (McKown, 2017). Although there are several popular student self-report SEL measures, few are aligned with the CASEL framework, and fewer still are well-adapted for applications within multitiered systems of support (MTSS). For example, CASEL recently conducted a review of SEL-focused assessments, including 16 self-report measures (CASEL, 2019). Of the reviewed assessments, most either do not assess more than three CASEL domains (rendering them

 Christopher J. Anthony, School of Special Education, School Psychology, and Early Childhood Studies, University of Florida;  Stephen N. Elliott, School of Social and Family Dynamics, Arizona State University;  James C. DiPerna and  Pui-Wa Lei, Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University.

Correspondence concerning this article should be addressed to Christopher J. Anthony, School of Special Education, School Psychology, and Early Childhood Studies, University of Florida, 2-189 Norman Hall, Gainesville, FL 32611. E-mail: canthony@coe.ufl.edu

Table 1
 Definitions of CASEL Social Emotional Learning Domains and Example Items From the SSIS SELb-S

Domain	Definition
Self-Awareness	The ability to accurately recognize one's emotions and thoughts, and their influence on behavior. This includes accurately assessing one's strengths and limitations and possessing a well-grounded sense of confidence and optimism. <i>Example item: I ask for help when I need it.</i>
Self-Management	The ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, controlling impulses, motivating oneself, and setting and working toward achieving personal and academic goals. <i>Example item: I stay calm when dealing with problems.</i>
Social-Awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports. <i>Example item: I help my friends when they are having a problem.</i>
Relationship skills	The ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups. This includes communicating clearly, listening actively, cooperating, resisting inappropriate social pressure, negotiating conflict constructively, and seeking and offering help when needed. <i>Example item: I try to forgive others when they say "sorry."</i>
Responsible decision-making	The ability to make constructive and respectful choices about personal behavior and social interactions based on consideration of ethical standards, safety concerns, social norms, the realistic evaluation of consequences of various actions, and the well-being of self and others. <i>Example item: I am careful when I use things that aren't mine.</i>

Note. CASEL = Collaborative for Academic Social and Emotional Learning; SSIS = Social Skills Improvement System; SELb-S = Social and Emotional Learning Brief Scales—Student Form. All definitions from CASEL (2012, p. 9).

nonoptimal for CASEL-focused assessment) or assess CASEL domains alongside many other tangentially related domains (rendering them inefficient for schools specifically using the CASEL framework). Only two reviewed assessments (the Washoe County School District Student Social Emotional Competency Assessment Crowder, Gordon, Brown, Davidson, & Domitrovich, 2019; and the Social Skills Improvement System—Social and Emotional Learning Edition Rating Form—Student; SSIS SEL-RF-S; Gresham & Elliott, 2017) focus on three or more CASEL domains without including many other tangential domains. Such limited assessment options are not unique to CASEL-aligned assessments (e.g., the National Center on Intensive Intervention [NCII] only lists one socioemotional or behavioral measure—the teacher-rated Social, Academic, and Emotional Behavior Risk Screener—as meeting psychometric standards for universal screening; NCII, 2019) but are particularly salient considering the growing prominence of this SEL model.

Of these two identified measures, only the SSIS SEL-RF-S was developed from a national sample, has national norms, and is aligned with a universal intervention program (the SSIS SEL Classwide Intervention Program; Elliott & Gresham, 2017). Despite these advantages, the SSIS SEL-RF-S is still relatively long (46 items requiring 10–12 min to complete). This length may hinder use within current MTSS models which emphasize large-scale (e.g., universal assessment/screening) or repeated (e.g., periodic progress monitoring) assessment. Given these considerations, a brief form of the SSIS SEL-RF-S would facilitate inclusion of the unique and, in some cases, more valid perspectives of students into MTSS assessment.

Rationale and Goals

In response to this need, we used item response theory (IRT) to develop the SSIS Social and Emotional Learning Brief Scales—

Student Form (SSIS SELb-S). In creating the SSIS SELb-S, we aimed to (a) significantly reduce the length of the SSIS SEL RF-S so it would take less than 5 min for students to complete (based on the estimated completion time of the SSIS SEL-RF-S, our goal was to include 20 items on the SSIS SELb-S), (b) retain appropriate content coverage and alignment to the CASEL competencies, (c) ensure a third grade (or lower) readability level, (d) produce scales yielding scores with sufficient reliability for low-stakes decisions, and (e) produce scales yielding scores with strong evidence of validity.

Method

Participants

Participants included all 800 students (8–18 years old) from the SSIS SEL-RF-S standardization sample. Demographic characteristics for this sample are reported in Table 2.

Measures

SSIS Social Emotional Learning Edition Rating Form—Student. The SSIS SEL RF-S (Gresham & Elliott, 2017) is a nationally normed behavior rating scale of social and emotional learning for students ages 8 to 18. The SSIS SEL RF-S includes 46 items rated on a 4-point scale from 0 (*Not True*) to 3 (*Very True*). With regard to reliability, Cronbach's alpha's for students ages 5 to 12 ranged from .73 to .95 (*Median* = .84) across the five SSIS SEL RF-S scales and the SEL composite. For students ages 13 to 18, Cronbach's alpha's ranged from .83 to .95 (*Median* = .89). Furthermore, 2-month stability coefficients for students between the ages of 8 and 18 were in the low .80s. Substantial evidence also exists for the validity of SSIS SEL RF-S scores. For example, the SSIS SEL RF-S Composite score was moderately negatively correlated with the BASC-2 (Reynolds & Kamphaus, 2004) Internalizing

Table 2
Demographic Characteristics of Participants (Percentages)

Characteristic	Sample ($N = 800$)	Population ^a
Female	50	49 ^b
Race/Ethnicity		
White	59	48
Black	15	15
Hispanic	19	26
Other	7	10
Grade		
3–5	41	30
6–8	36	23
9–12	23	15
Region		
Northeast	18	16
Midwest	22	21
South	36	39
West	24	24
Parent's education level		
Grade 11 or less	13	11
Grade 12 or GED	29	19
1–3 years of college	30	26
4+ years of college	27	44
Educational status		
General education	91	87
Special education	9	13

Note. Some percentages do not sum to 100 due to rounding.

^a Except where noted, estimates derived from the 2016–2017 Digest of Educational Statistics (Snyder, de Brey, & Dillow, 2018). ^b Derived from the most recent estimates from the 2013–2014 Civil Rights Data Collection survey (U.S. Department of Education Office for Civil Rights, 2016).

Problems Composite ($r = -.28$), School Problems Composite ($r = -.43$), Inattention/Hyperactivity Composite ($r = -.53$), and Emotional Symptoms Index ($r = -.40$), but moderately positively correlated with the Personal Adjustment Composite ($r = .50$). Confirmatory factor analyses also provided support of the internal structure of the SSIS SEL RF-S yielding a five-factor model (Gresham et al., 2018).

Behavior Assessment System for Children-Second Edition Self Report of Personality (BASC-2). The BASC-2 (Reynolds & Kamphaus, 2004) is a norm referenced diagnostic tool for assessing the behavior of children and youth. The BASC-2 Self-Report Child (ages 8–12) form consists of 139 items that represent 14 primary scales and five composite scales (Emotional Symptoms Index, Inattention/Hyperactivity, Internalizing Problems, Personal Adjustment, and School Problems); the Adolescent (ages 13–18) form consists of 176 items that represent 16 primary scales that also contribute to the same five composite scales. The BASC-2 manual reports internal consistency coefficients in the .80s for the composite scales for both the Child (Range = .72–.86) and Adolescent (Range = .67–.88) forms. Test–retest reliability yielded high correlations for composite scores for both the Child (Range = .64–.82) and Adolescent (Range = .63–.84) forms. The BASC-2 Technical Manual also provides a number of concurrent comparisons of composite scores for the Child and Adolescent forms with established measures of childhood behavior (e.g., ASEBA Youth Self-Report), which are generally supportive of the validity of BASC-2 scores.

Procedure

Data used in the current study were collected as part of the original SSIS Rating Scale standardization. Pearson Assessment field staff recruited school site coordinators in 115 schools across 36 states. These site supervisors then recruited participants to fit demographic targets based on the 2006 Current Population Survey. These site coordinators and their schools distributed and collected the rating scales from September 2006 to October 2007. The final standardization sample was selected from the larger respondent sample to fit 2006 U.S. Census¹ demographics of age, sex, race/ethnicity, and educational status. A subsample of 127 students rated themselves twice for test–retest reliability calculation. Finally, a subsample of 53 students also completed the BASC-2.

Data Analysis

Based on the factor analytic work of Gresham et al. (2018) substantiating the five-factor structure of the SSIS SEL-RF-S and in line with our goal of developing a CASEL-aligned assessment, we conducted IRT analyses by subscale to identify items for the SSIS SELb-S. Prior to IRT analyses, we checked key IRT assumptions. First, we examined unidimensionality by conducting exploratory factor analyses on each SSIS SEL RF-S subscale with MPlus (Muthén & Muthén, 2019) using the essential unidimensionality criterion that the ratio of first to second eigenvalues should exceed 4 (Reeve et al., 2007). We also conducted parallel analysis for ordered polytomous data on each SSIS SEL RF-S scale (Timmerman & Lorenzo-Seva, 2011) with the FACTOR program (Lorenzo-Seva & Ferrando, 2006) to ensure no more than one observed eigenvalue was above the chance level. Next, we evaluated the assumption of local independence using standardized local dependence χ^2 values output by IRTPRO with a recommended cutoff value of 10 (Cai, Thissen, & du Toit, 2019) and addressed local independence violations in item selection. To evaluate overall model fit, we focused on the RMSEA values output by IRTPRO (Cai et al., 2019) with a cutoff value of $\leq .06$ indicating acceptable fit (Hu & Bentler, 1999). Once assumptions had been checked, we used the Graded Response Model (Samejima, 1969) for IRT analyses in line with previous investigations (e.g., Anthony, DiPerna, & Lei, 2016). Because prior research has supported applications of this model with samples as small as 500, we considered our sample size of 800 sufficient (Forero & Maydeu-Olivares, 2009).

We then selected items for the SSIS SELb-S based on several indices as well as item content. First, we used item information curves with a special focus on the “at risk” range (-1.5 to -0.5 on the θ scale; Anthony et al., 2016). This range was chosen given the anticipated use of the SSIS SELb-S with some students with mild to moderate SEL difficulties. We considered item selection with the goal of achieving a .70 reliability in this range. Second, as mentioned above, we addressed local dependence when selecting items. Third, we conducted differential item functioning (DIF) analyses for each SSIS SEL RF-S item for both sex (male vs. female) and race/ethnicity (white vs. nonwhite). We used the two-step DIF procedure discussed by Tay, Meade, and Cao (2015),

¹ This sample was also subsequently confirmed to represent the 2014 U.S. Census as well during the SSIS SEL development.

and, if DIF was found to be present, we calculated the expected score standardized difference (ESSD) between focal and reference groups using the Visual DF program (Meade, 2010). This index is in standard deviation units and can thus be interpreted similarly to Cohen's d (0.2 = small; 0.5 = medium; 0.8 = large; Cohen, 1988). With these sources of information, each of the authors independently reviewed and consulted CASEL definitions for each content domain and selected 4 to 5 items per scale for the SSIS SEL*b*-S. Initial item selection agreement across the authors generally ranged from 50% to 75% depending on the scale. Thereafter, all four authors discussed item options based on the psychometric data and content representation until reaching consensus on final item selections.

After identifying the SSIS SEL*b*-S, we conducted reliability analyses on these sets of items. First, we examined test information functions (TIFs) for each identified SSIS SEL*b*-S scale. Second, we computed Cronbach's alpha for each SSIS SEL*b*-S scale and the SSIS SEL*b*-S SEL composite.² Third, we used the test-retest sample of the SSIS SEL RF-S standardization sample ($n = 127$) to compute test-retest reliability coefficients. Fourth, we conducted initial validity analyses on the SSIS SEL*b*-S by computing interscale correlations for all SSIS SEL*b*-S scales and the SSIS SEL*b*-S SEL composite. We also calculated correlations between the SSIS SEL*b*-S scales with the BASC-2. For comparative purposes, we also computed equivalent reliability and validity statistics with the SSIS SEL RF-S.

Results

As noted previously, we initially checked IRT assumptions for the SSIS SEL RF-S scales. Our criterion for essential unidimensionality was met for each scale, with ratios of first to second eigenvalues ranging from 4.09 to 5.05 (*Median* = 4.63). Parallel analysis also supported unidimensionality in all cases. The percentage of item pairs evidencing local dependence ranged from 0% to 36% (*Median* = 3%) across scales. During item selection, most of these violations were addressed, and only two items with local dependence were retained on the SSIS SEL*b*-S. The degree of local dependence violation between these items only slightly exceeded our a priori criterion (standardized $\chi^2 = 10.8$) and both were from the SSIS SEL RF-S Self-Management scale, which had the most items with local item dependence and thus the fewest alternative options. Also, RMSEA values were acceptable in all cases with values of .03 for the SSIS SEL RF-S Social Awareness scale, .06 for the SSIS SEL RF-S Self-Management scale, and .04 for all remaining SSIS SEL RF-S scales. With regard to DIF violations, the number of items exhibiting sex DIF ranged from 0 to 2 (*Median* = 1) across SSIS SEL RF-S scales, and the number of items exhibiting race/ethnicity DIF ranged from 0 to 3 (*Median* = 2). As a result, no items with race/ethnicity DIF were retained on the SSIS SEL*b*-S, and only two items exhibiting sex DIF were retained. These retained items had ESSD values of .48 (*I try to find a good way to end a disagreement*) and .37 (*I follow school rules*) respectively indicating small to medium overall DIF effects in favor of females. Considering these statistical indices and item content, 20 items (4 per scale) were selected for the SSIS SEL*b*-S. Example items are provided in Table 1.

As shown in Figure 1, SSIS SEL*b*-S scales maintained a .70 level of reliability across broad levels of each SEL construct. With

regard to traditional reliability evidence, Cronbach's alpha levels were .91 for the SSIS SEL*b*-S SEL Composite and ranged from .67 to .72 (*Median* = .69) across SSIS SEL*b*-S scales. Test-retest reliability coefficients were .87 for the SSIS SEL*b*-S SEL Composite and ranged from .64 to .83 (*Median* = .71) across SSIS SEL*b*-S scales. With regard to validity evidence, SSIS SEL*b*-S interscale correlations were moderate (.55 - .65), and correlations with the BASC-2 followed expected patterns (see Table 3). SSIS SEL*b*-S scales generally demonstrated small to moderate positive correlations with the BASC-2 Personal Adjustment Composite and small to moderate negative correlations with the BASC-2 School Problems, Internalizing Problems, Inattention/Hyperactivity, and Emotional Symptoms composites. Notable exceptions included the correlation between the SSIS SEL*b*-S Self-Awareness scale and the BASC-2 Internalizing Problem Composite ($r = -.02$), the correlation between the SSIS SEL*b*-S Social Awareness scale and the BASC-2 Internalizing Problems Composite, ($r = .01$), and the correlation between the SSIS SEL*b*-S Social Awareness scale and the BASC-2 Emotional Symptoms Index ($r = -.09$).

Discussion

The primary outcome of this study was the initial development and validation of the SSIS SEL*b*-S, a brief self-report measure of students' SEL skills aligned with the CASEL SEL framework. Overall, evidence from the study provides support for the reliability and validity of SSIS SEL*b*-S scores. Reliability evidence generally met common thresholds for low-stakes decision making (Salvia, Ysseldyke, & Witmer, 2017), and validity evidence provided initial support for the intended uses of SSIS SEL*b*-S scores. Furthermore, in line with our goals, the SSIS-SEL*b*-S has a Flesch-Kincaid readability level of 2.5. Although additional studies are necessary, the current evidence suggests that the SSIS-SEL*b*-S demonstrates promise for identifying SEL strengths and weaknesses to inform intervention planning. The SSIS SEL*b*-S might also be useful as a universal screener or periodic progress monitoring measure to assess response to evidence-based interventions aligned with the CASEL SEL framework such as the SSIS SEL-CIP (Elliott & Gresham, 2017). Although the length of the SSIS SEL*b*-S is appropriate for such applications, it is important to note that further evidence specifically supporting these uses will be necessary (e.g., classification accuracy data to support universal screening). Future research should also examine aspects of the usability of the SSIS SEL*b*-S including evaluating administration time, decision rules, and the social validity of the SSIS SEL*b*-S.

Although the findings are generally promising, there are several important limitations to consider. First, though sufficient for low-stakes decision making (Salvia et al., 2017), the achieved reliability for the SSIS-SEL*b*-S is lower than other attempts at creating brief scales using IRT (e.g., Anthony et al., 2016). It is important to note, however, that none of these previous efforts focused on student self-report forms, which tend to produce scores with lower precision (e.g., Jenkins et al., 2014). In addition, evidence from TIFs indicated that reliability is lowest at higher levels of each

² Because student self-report measures tend to be less reliable than teacher and parent reports (e.g., Jenkins et al., 2014), we used the .70 reliability criterion suggested by Salvia et al. (2017) for progress monitoring decisions as our goal during development.

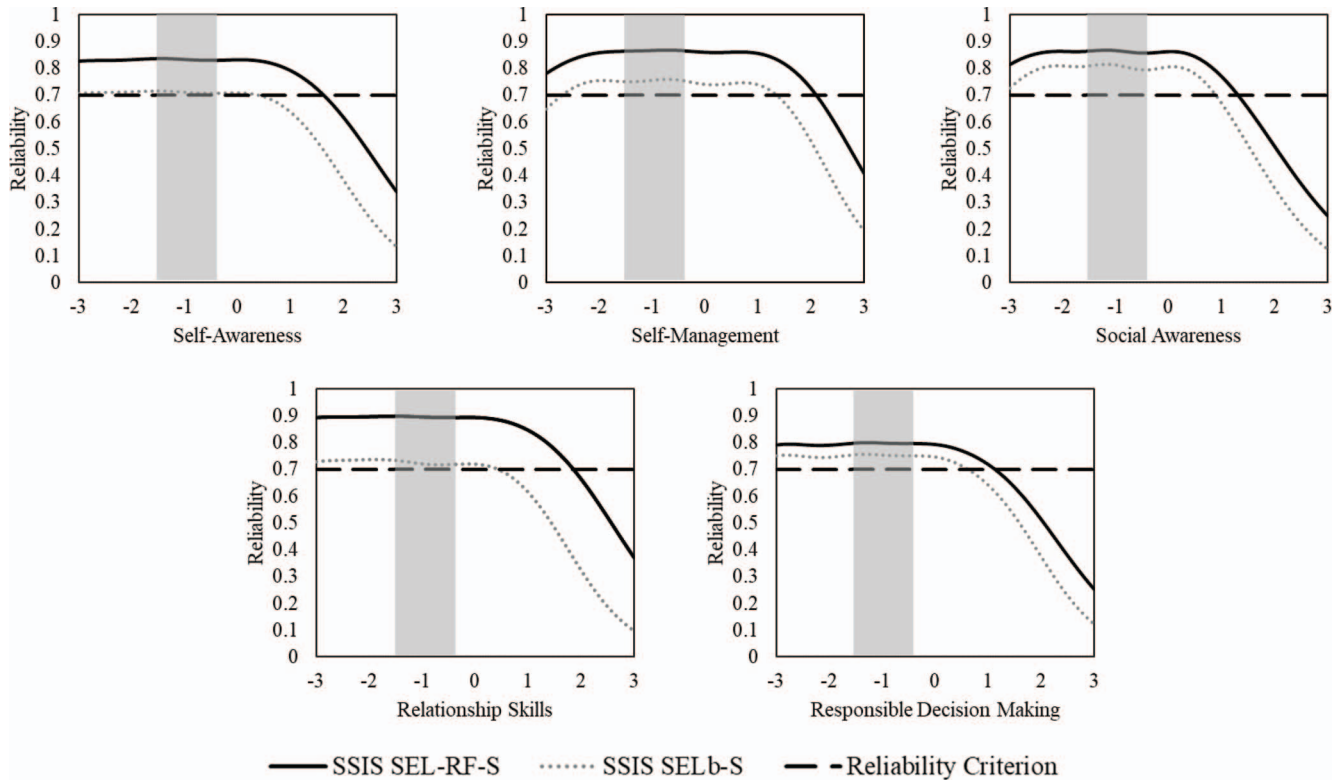


Figure 1. Test Information Functions for SSIS SEL Rating Form–Student and SSIS SEL Brief Scales–Student Form. Shaded region represents “at risk” range. Total information on y-axis converted to reliability metric with the following formula: $1 - (1/\text{information})$ as recommended by Petrillo, Cano, McLeod, and Coon (2015).

SEL domain, likely due to a ceiling effect. Thus, users should be aware that high scores on the SSIS SELb-S are less precise than lower scores. Also, we retained several items with evidence of sex DIF. These differences are in line with longstanding research indicating meaningful sex-differences in social behavior (Del Giudice, 2015). Yet, because the magnitude of DIF was small to moderate and these items were on different SSIS SELb-S scales, their inclusion is unlikely to result in scale-level differential functioning.³

In addition, although findings from Gresham et al. (2018) supported the 5-factor structure of the SSIS SEL-RF-S, it is possible that other factor structures might also plausibly characterize the SSIS SEL-RF-S. This is particularly salient because the SSIS SEL-RF-S represents a reconfiguration of items from a different measure, the Social Skills Improvement System—Rating Form—Student (Gresham & Elliott, 2017), and we used the sample that served as the standardization sample for both measures. However, in addition to statistical considerations, dimensionality decisions should be guided by substantive and practical considerations (Ackerman, Gierl, & Walker, 2003; Tate, 2002) and these latter factors clearly favor the 5-factor structure aligned with the prominent CASEL framework (Eklund et al., 2018).

An additional consideration is that the sample and sole validity measure (the BASC-2) used in this study are over 10 years old. Although it is often assumed that psychoeducational measures remain current for roughly 10 years (e.g., Kamphaus, 2005), this and similar rules of thumb are typically made in reference to

standardized intelligence and achievement tests for which there is a known mechanism that renders norms obsolete (i.e., the Flynn effect). There currently is no similar empirical evidence that social and emotional behavior changes in such a manner and thus, it is unclear if the 10-year recommendation is appropriate in this domain. Future research should evaluate whether and how social and emotional behavior changes on a population level to better substantiate such considerations. Finally, although the BASC-2 measures emotional and behavioral constructs related to SEL skills, other constructs such as social skills and emotion recognition are likely more closely linked with SEL than most BASC-2 constructs. Thus, future validation research using measures of these more closely related constructs will be important to provide further relational evidence for the SSIS SELb-S.

Although there are a number of student SEL measures, few are well-aligned with the CASEL framework. Fewer still are brief enough to be feasible for many MTSS applications such as universal screening and periodic progress monitoring. By aligning with this framework and emphasizing efficiency, the SSIS SELb-S is unique and holds promise for involving students in the process of assessing and improving SEL competencies. Such alignment also facilitates integration of information with other CASEL-aligned measures for different informants (e.g., teachers, parents)

³ If users are concerned about this level of DIF, the use of sex-specific normative scores would also address this issue.

Table 3

SSIS Brief SEL Student and SSIS-SEL-Student Form Correlations With the BASC-2

BASC-2	Self-Awareness		Self-Management		Social-Awareness		Relationship skills		Responsible decision-making		SEL composite	
	SEL <i>b</i>	SEL	SEL <i>b</i>	SEL	SEL <i>b</i>	SEL	SEL <i>b</i>	SEL	SEL <i>b</i>	SEL	SEL <i>b</i>	SEL
School problems	-.22	-.35	-.50	-.45	-.20	-.34	-.34	-.37	-.29	-.33	-.38	-.44
Internalizing problems	-.02	-.16	-.33	-.30	-.01	-.06	-.33	-.25	-.30	-.28	-.22	-.25
Inattention/Hyperactivity	-.51	-.54	-.48	-.49	-.24	-.32	-.42	-.45	-.54	-.54	-.53	-.54
Emotional symptoms	-.14	-.29	-.41	-.36	-.09	-.14	-.44	-.40	-.42	-.39	-.36	-.38
Personal adjustment	.34	.42	.44	.36	.28	.31	.51	.52	.53	.47	.51	.49

Note. SEL*b* = Social Skills Improvement System Social and Emotional Learning Edition Brief Scales–Student; SEL = Social Skills Improvement System Social and Emotional Learning Edition Rating Form–Student; BASC-2 = Behavior Assessment System for Children–Second Edition. Bolded coefficients indicate statistically significant differences between validity coefficients calculated using Steiger's (1980) formula.

and intervention programming. Considering the importance of students' voice in the SEL movement, this expansion and integration represents a positive next step toward more comprehensive assessment for effective SEL programming in schools.

References

- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). An NCME instructional module on using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22, 37–51. <http://dx.doi.org/10.1111/j.1745-3992.2003.tb00136.x>
- 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. <http://dx.doi.org/10.1016/j.jsp.2015.12.005>
- Cai, L., Thissen, D., & du Toit, S. H. C. (2019). IRTPRO for Windows (Version 2.1) [Computer software]. Lincolnwood, IL: Scientific Software International.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Collaborative for Academic, Social, and Emotional Learning. (2012). *2013 CASEL guide: Effective social and emotional learning programs—Preschool and elementary school ed.* Chicago, IL: Author.
- Collaborative for Academic Social and Emotional Learning. (2019). CASEL assessment guide. Retrieved from <https://measuringcel.casel.org/access-assessment-guide/>
- Crowder, M. K., Gordon, R. A., Brown, R. D., Davidson, L. A., & Domitrovich, C. E. (2019). Linking social and emotional learning standards to the WCSD Social-Emotional Competency Assessment: A Rasch approach. *School Psychology*, 34, 281–295. <http://dx.doi.org/10.1037/spq0000308>
- Del Giudice, M. (2015). Gender differences in personality and social behavior. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 750–756). New York, NY: Elsevier. <http://dx.doi.org/10.1016/B978-0-08-097086-8.25100-3>
- Eklund, K., Kilpatrick, K. D., Kilgus, S. P., & Haider, A. (2018). A systematic review of state-level social–emotional learning standards: Implications for practice and research. *School Psychology Review*, 47, 316–326. <http://dx.doi.org/10.17105/SPR-2017.0116.V47-3>
- Elliott, S. N., & Gresham, F. M. (2017). *Social Skills Improvement System Social-Emotional Learning Edition: Classwide Intervention Program*. Minneapolis, MN: Pearson Assessments.
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, 14, 275–299. <http://dx.doi.org/10.1037/a0015825>
- Gresham, F. M., & Elliott, S. N. (2017). *Social Skills Improvement System Social Emotional Learning Edition Rating Forms*. Bloomington, MN: Pearson Assessments.
- Gresham, F. M., Elliott, S. N., Metallo, S., Byrd, S., Erickson, M., & Altman, R. A. (2018). Psychometric fundamentals of the SSIS SEL Rating Forms. *Assessment for Effective Intervention*, 45, 1–16. <http://dx.doi.org/10.1177/1534508418808598>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., . . . Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary schools. *Contemporary School Psychology*, 18, 241–254.
- Kamphaus, R. W. (2005). *Clinical assessment of child and adolescent intelligence* (2nd ed.). New York, NY: Springer. <http://dx.doi.org/10.1007/978-0-387-29149-9>
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38, 88–91. <http://dx.doi.org/10.3758/BF03192753>
- McKown, C. (2017). Social-emotional assessment, performance, and standards. *The Future of Children*, 27, 157–178. <http://dx.doi.org/10.1353/foc.2017.0008>
- Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *Journal of Applied Psychology*, 95, 728–743. <http://dx.doi.org/10.1037/a0018966>
- Muthén, L. K., & Muthén, B. (2019). *Mplus. The comprehensive modelling program for applied researchers: User's guide*, 5. Los Angeles, CA: Author.
- National Center on Intensive Intervention. (2019). Behavior screening tools chart. Retrieved from <https://charts.intensiveintervention.org/chart/behavior-screening>
- Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, 18, 25–34. <http://dx.doi.org/10.1016/j.jval.2014.10.005>
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., . . . the PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45(Suppl. 1), S22–S31. <http://dx.doi.org/10.1097/01.mlr.0000250483.85507.04>

- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Salvia, J., Ysseldyke, J. E., & Witmer, S. (2017). *Assessment in special and inclusive education* (13th ed.). Boston, MA: Cengage Learning.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 34(Suppl.), 1–97. <http://dx.doi.org/10.1007/BF03372160>
- Smith, S. R. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment*, 25, 139–149. <http://dx.doi.org/10.1177/0734282906296233>
- Snyder, T. D., de Brey, C., & Dillow, S. A. (2018). *Digest of education statistics 2016* (NCES 2017–094). Washington, DC: NCES, IES, U.S.D.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251. <http://dx.doi.org/10.1037/0033-2909.87.2.245>
- Tate, R. L. (2002). Test dimensionality. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 181–211). Mahwah, NJ: Erlbaum.
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18, 3–46. <http://dx.doi.org/10.1177/1094428114553062>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16, 209–220. <http://dx.doi.org/10.1037/a0023353>
- U.S. Department of Education Office for Civil Rights. (2016). *Civil rights data collection: A first look*. Retrieved from <https://www2.ed.gov/>
- Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York, NY Guilford Press.

Received November 21, 2019

Revision received May 8, 2020

Accepted May 15, 2020 ■

Correction to Kilgus et al. (2019)

In the article “Developing a Direct Rating Behavior Scale for Depression in Middle School Students,” by Stephen P. Kilgus, Michael P. Van Wie, James S. Sinclair, T. Chris Riley-Tillman, Keith C. Herman, *School Psychology*, 2019, Vol. 34, No. 1, pp. 86–95. <http://dx.doi.org/10.1037/spq0000263>, the error occurred in the Author’s Note which omitted the funder information. It should read: “The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130143 to the University of Missouri (PI: Keith Herman). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.” The online version of this article has been corrected.

<http://dx.doi.org/10.1037/spq0000386>