

Screening the Whole Social-Emotional Child: Expanding a Brief SEL Assessment to Include Emotional Behavior Concerns

Stephen Nelson Elliott , Pui-Wa Lei , Christopher. J. Anthony & James Clyde DiPerna

To cite this article: Stephen Nelson Elliott , Pui-Wa Lei , Christopher. J. Anthony & James Clyde DiPerna (2021): Screening the Whole Social-Emotional Child: Expanding a Brief SEL Assessment to Include Emotional Behavior Concerns, School Psychology Review, DOI: [10.1080/2372966X.2020.1857659](https://doi.org/10.1080/2372966X.2020.1857659)

To link to this article: <https://doi.org/10.1080/2372966X.2020.1857659>



Published online: 29 Jan 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Screening the Whole Social-Emotional Child: Expanding a Brief SEL Assessment to Include Emotional Behavior Concerns

Stephen Nelson Elliott^a, Pui-Wa Lei^b , Christopher. J. Anthony^c, and James Clyde DiPerna^b

^aArizona State University; ^bPennsylvania State University; ^cUniversity of Florida

ABSTRACT

Universal screening of students is increasing as programs are implemented for improving social emotional learning (SEL) skills. Assessments used to conduct SEL screenings focus on social emotional strengths; however, such assessments can provide a broader characterization of students' wellbeing by concurrently screening for emotional behavior concerns (EBC). This article describes the development of brief EBC-Internalizing and EBC-Externalizing scales designed to augment the SSIS SEL Brief Scales (SSIS SELb) Teacher and Student Forms. To further evaluate the utility of concurrently screening for SEL and EBCs, we examined distributions of SEL scores at each EBC Concern level. Both the Teacher and Student assessments of SEL skills and EBC behaviors maximized efficiency and yielded reliable and valid scores that provide insights regarding the interplay of these behavior constructs. The SSIS SELb and EBC scales can be administered together as the SSIS SELb + Mental Health Scales to provide efficient measurement of the whole social emotional child. The article examines the study's limitations, follow-up research, and implications for using brief universal screenings and instructional programs to advance the wellbeing of students.

IMPACT STATEMENT

Universal screening of children's social emotional learning (SEL) competencies has been strength focused. A meaningful percentage of students identified as functioning at competent or advanced SEL levels, however, concurrently report cooccurring internalizing or externalizing emotional behavior concerns (EBC). Thus, universal screening programs that integrate EBC and SEL measurement can better serve the whole social emotional child.

ARTICLE HISTORY

Received June 25, 2020
Accepted November 25, 2020

KEYWORDS

mental health screening, universal screening, social emotional learning, emotional behavior concerns, criterion-referenced scores, whole social emotional child assessment

ASSOCIATE EDITOR

Tyler Renshaw

Programs for improving students' social emotional wellbeing are prevalent in schools as the result of research showing social emotional learning (SEL) competencies influence students' academic engagement and achievement (e.g., DiPerna et al., 2002, 2005, 2016, 2018) and decrease some concurrent problem behaviors (e.g., DiPerna et al., 2015). The Collaborative for Academic, Social and Emotional Learning (CASEL) also has influenced the increase in SEL programs through its advocacy collaborations with states to develop content standards for students' SEL competencies (Dusenbury et al., 2019). Virtually all of these school-based SEL programs and content standards are based on the strength-focused CASEL competency framework (<http://www.CASEL.org>).

SEL Competencies: A Strength-Focused Perspective

The CASEL Competency Framework was derived by Utne O'Brien and Weissberg (CASEL, 2003) and has been empirically supported with adolescents (Ross & Tolan,

2017) as well as a national representative sample of children ages 3–18 (Gresham et al., 2018). Many of the SEL competencies that society values have been articulated in this competency framework. Specifically, it emphasizes the intra-personal competencies of self-awareness and self-management, the inter-personal competencies of social awareness and relationship skills, and a fifth domain, responsible decision making, that is considered both an inter- and intra-personal competency. These competency domains, and the many skills representative of them, can be assessed, taught, and improved, thus advancing children's chances of functioning well and preventing social emotional problems at school, home, and in their communities.

The strength-focused conceptualization of social emotional behavior advanced by CASEL is positive, appealing, and popular in the United States; however, by definition it is narrower and unbalanced when the range of possible social emotional behaviors that can affect children's wellbeing are considered. For example, notably absent in most of the SEL programs—as well as virtually all assessments

that align with the CASEL Competency Framework—is consideration of students’ emotional behavior concerns (EBC). From the perspective of the CASEL logic model, emotional behavior concerns such as internalizing or externalizing problems would represent an outcome of suboptimal SEL skill development. Indeed, there is ample evidence that SEL skills promote lower incidence of emotional or behavioral problems (e.g., DiPerna et al., 2015; Durlak et al., 2011). Yet, such evidence is probabilistic rather than deterministic and as such, assuming that all children who exhibit low SEL skills will also exhibit EBCs or that all children who exhibit high SEL skills will not experience EBCs is speculative. Common EBCs or undesirable behaviors include internalizing concerns (e.g., anxiousness, depression) and externalizing concerns (e.g., aggressiveness, bullying). These behavior concerns and outcomes understandably do not fit the strength-focus, positive psychology orientation guiding most SEL programs.

A significant number of students, perhaps 10% to 20%, will have difficulty learning and applying SEL skills and concurrently display some undesirable or negative emotional behaviors that often interfere with their response to interventions focusing on desired social skills (e.g., Forness et al., 2012; Saeki et al., 2011). In addition, many of these EBCs can be assessed along with desired SEL behaviors and some likely will be reduced in SEL programs when students increase their use of positive behaviors taught in SEL programs or related social skills programs (Elliott & Gresham, 2008, 2020). Thus, creating and using an integrated SEL and mental health screening assessment with scores based on the same standardization population, rather than separate universal SEL screening and separate mental health screening programs within the same schools, would likely require less administration time, yield an integrated report that provides a more holistic characterization of students’ social emotional functioning, and may enhance early intervention support. Furthermore, recent evidence indicates that few schools (roughly 12%) conduct any socioemotional or behavioral universal screening at all (Bruhn et al., 2014). For the remaining 88% of schools, even completing a single screener apparently is burdensome. Thus, the prospect of completing, integrating, and interpreting more than one screener representing multiple socioemotional and behavioral domains of interest would be very unlikely. For these schools, a single, efficient measure integrating results from key domains would substantially support the goals of comprehensive universal screening. Such a value-added argument provided the main rationale for the development of an integrated universal screening assessment of CASEL aligned SEL competencies and key EBC behaviors.

The Whole Social Emotional Child and Existing Assessments

Current models of mental health conceptualize complete mental health as being composed of two distinct dimensions: one dimension involves psychosocial wellbeing/positive experiences and the other dimension involves psychosocial distress/negative experiences (e.g., National Institute of Mental Health, 2020; Petersen et al., 2020). Although one may be primarily interested in improving students’ specific SEL skills and overall competencies, it can be efficient and informative to concurrently screen and monitor students for behavioral indicators of psychological distress as part of multitiered screening assessments and interventions. Thus, it is important to consider advantages (time, risk management, intervention planning, costs, etc.) of a broadband (SEL skills + emotional behavior concerns) versus an SEL skills only screening assessment of students relative to SEL program goals that stress strength focused, “whole” child intervention. Furthermore, explicitly assessing EBCs addresses the assumptions that all students who exhibit low SEL skills will also experience EBCs and that all students who exhibit high SEL will not experience EBCs. This assumption is problematic when applied to individual cases and addressed by explicitly assessing EBC behaviors alongside SEL skills.

To increase the likelihood of SEL programs addressing the needs of the whole social emotional child, universal screening assessments are needed that document the range of both positive and negative social emotional behaviors commonly observed in K-12 students. Ideally, such assessments efficiently measure all the positive behaviors taught in the SEL program, as well as a sample of negative emotional behavior concerns that might interfere with the effective production of SEL skills, thus providing a more balanced characterization of students’ wellbeing. A focus on both positive and negative social emotional behaviors of children is consistent with a dual-factor approach to conceptualizing wellbeing or mental health (e.g., Greenspoon & Saklofske, 2001; Petersen et al., 2020) and the measurement history of the Social Skills Rating System (Gresham & Elliott, 1990) and its successor, the Social Skills Improvement System Rating Scales (Gresham & Elliott, 2008). Both multiinformant behavior rating scales were developed to not only assess students’ social skills (e.g., cooperation, engagement, empathy, self-control), but also concurrent problem behaviors (e.g., internalizing, externalizing, hyperactive/inattention) and academic competencies to provide point-in-time characterization of students’ social and academic functioning. Assessments like the SSIS Rating Scales and the Behavior Assessment Scales for Children-3 (Reynolds & Kamphaus, 2015) are

both broadband measures, with the former focusing predominately on the positive social and emotional side of children's wellbeing and the latter focusing on the negative or problem behavior side of children's wellbeing. Both are technically sound assessments, but too long for use in screening (especially if used simultaneously), and neither were designed specifically to align with the CASEL Competency Framework.

Not surprisingly, no CASEL competency aligned, time efficient (10 min or less) SEL screening measures are available for conducting universal screenings of the whole social emotional child. At present, schools primarily use teacher referrals, rather than formal assessments, to identify students at risk for social, emotional, and behavioral difficulties (e.g., Bruhn et al., 2014). Although teachers generally identify students with externalizing difficulties accurately, they routinely have been found to under-identify students experiencing internalizing problems (e.g., Dowdy et al., 2013). Teachers in middle and high schools, compared to elementary teachers, also have more difficulty identifying students with internalizing concerns because they generally have less opportunity to interact with their students.

Assessments of students' SEL competencies predominately have been behavior rating scales completed by teachers and students, although a few rating scales have been designed for parents or family members to complete (Denham, 2015). Rating scales, when well designed, can be time-efficient and yield reliable and valid scores for their intended uses.

Teacher SEL Assessments

The CASEL Assessment Guide identifies eight teacher/staff report measures (CASEL, 2019). Of these assessments, only one is sufficiently brief for universal screening, the Devereux Student Strengths Assessment—Mini (DESSA-mini; Naglieri et al., 2011/2014). This measure consists of eight items representing the five CASEL competencies plus some additional items representing related academic skills. The DESSA-mini is time efficient, aligned with the CASEL framework, yields norm-referenced scores, and has evidence for score reliability and validity. It, however, only provides information on students' SEL skills globally and thus not able to identify specific competency areas in need of improvement.

A new screening option is the SSIS SEL Brief Scales (SSIS SELb; Elliott et al., 2020a), which includes a Teacher Form. This scale consists of 20 items (four for each of the five CASEL competencies) and can be completed in 5 min. This assessment was developed using an Item Response Theory (IRT) approach to select maximally efficient items from the norm-referenced SSIS SEL Rating

Form—Teacher (Gresham & Elliott, 2017). Research indicates there is substantial evidence for the reliability and validity of the score from the SSIS SELb (Anthony et al., 2020c). Like the DESSA-Mini, the SSIS SELb only assesses SEL competencies; however, it does provide scores for each of the five CASEL competency domains to better enable the identification of areas in need of improvement, and it links to the SSIS SEL Class-wide Intervention Program, a CASEL recognized SElect program (2018).

Student SEL Assessments

Although there are several self-report SEL measures for students in Grades 3 and above, few are aligned with the CASEL framework, and fewer still are well-adapted for applications as large-scale screening assessments. For example, a review of SEL-focused assessments indicated there are 16 self-report measures (CASEL, 2020). Only two reviewed assessments, the Washoe County School District Student Social Emotional Competency Assessment (Crowder et al., 2019) and the SSIS SEL Edition Rating Form—Student (SSIS SEL-RF-S; Gresham & Elliott, 2017) focus on three or more CASEL competency domains without including many other non-SEL domains.

A Student form of the SSIS SELb Scales (Elliott et al., 2020a) also has been developed. Like the corresponding SSIS SELb Scales—Teacher, the Student version contains 20 SEL items representative of the five CASEL competencies. This assessment also is a shortened version of the SSIS SEL Rating Form-Student (Gresham & Elliott, 2017), can be completed in 5 min, and yields reliable and valid scores for universal screening of SEL skills. It, however, does not measure any emotional behavior concerns.

Purpose, Questions, and Research Strategy

Although there are efficient measures of students' skills representative of SEL competencies as defined by the CASEL model that can be used for universal screening of large groups of students, none concurrently targets EBC. This situation is likely the result of many involved in SEL programs wanting to prioritize the positive side of human behavior and not deal with "psychopathology or deficit" (e.g., McKown, 2019, p. xxii). This is problematic because though SEL skills and EBC are related, they are not synonymous. Only measuring SEL skills could lead to the under-identification of critically important student EBC. As such, the current study sought to expand the SSIS SELb Teacher and Student Forms through the development of EBC scales from the established Problem Behaviors Scale of the SSIS Rating Scales. Specifically, this study focused on the development of an EBC-Internalizing (EBC-I) scale and an EBC-Externalizing (EBC-E) scale from the larger

pool of the SSIS Problem Behaviors items. Item Response Theory (IRT) procedures like those used to develop the SSIS SELb (e.g., Anthony et al., 2016; Anthony & DiPerna, 2017, 2018; Moulton et al., 2019) were used to select the items. Subsequently, we examined the initial reliability and validity of scores from these scales and the utility of using them concurrently with the SSIS SELb Scales. Score reliability estimates were evaluated using internal consistency (Cronbach's α and omega hierarchical), test-retest, and interrater reliability coefficients. Several forms of validity evidence were examined including internal structure (confirmatory factor analyses, and intercorrelations between EBC-I and EBC-E Scales), and relationship with related constructs (correlations between the EBC Scales and SSIS SELb Scales, and correlations between the EBC Scales and the BASC-2).

The key questions addressed in the study were: (1) *Is there substantial evidence that the scores from the EBC scales are reliable and valid for universal screening?* and (2) *Does the direct assessment of emotional behavior concerns identify more students with emotional concerns than when only an assessment of SEL skills alone is used?* The research strategy used to provide a data-based answer to these motivating questions involved examining the scores on the developed EBC scales from a nationally representative sample of children in Grades K-12 to evaluate (a) the initial reliability and validity of EBC-I and EBC-E scores and (b) the utility of criterion-referenced scoring frameworks for describing students' SEL competencies and emotional behavior concerns. Together, this evidence was expected

to support claims that (a) the developed EBC scales yield scores with strong psychometric evidence to support their use and (b) using them concurrently with SEL-focused assessment provides practical insights regarding students' need for social emotional support within CASEL-aligned SEL programs.

METHOD

Participants

Participants were drawn from the standardization samples of the Teacher and Student SSIS-Rating Scales (SSIS-RS; Gresham & Elliott, 2008) and included 750 teachers and 800 student participants for the development of the EBC-Teacher Scales and EBC-Student Scales, respectively. The student participants were diverse across race/ethnicity, region, SES, and educational status (Table 1).

Measures

The measures for this study included three published behavior rating scales, the SSIS Rating Scale (Gresham & Elliott, 2007), the SSIS SELb Rating Scale (Elliott et al., 2020b), and the Behavior Assessment for Children Scale-2 (BASC-2; Reynolds & Kamphaus, 2004). The Problem Behavior items and standardization data from the original SSIS-RS were used as the basis for selecting EBC items. (Note that the SSIS SELb also is comprised of items from the SSIS-RS.) The SSIS SELb and BASC-2 were used as a measure for testing the concurrent relation of the new EBC

Table 1. Demographic Characteristics of Participants (Percentages)

Characteristic	Teacher Sample	Student Sample	Current U.S. Student Population ^a
Female	50	50	49 ^b
Race			
White	59	59	48
Black	16	15	15
Hispanic	19	19	26
Other	6	7	10
Grade			
K-2	32	—	22
3-5	27	41	23
6-8	25	36	22
9-12	16	23	30
Region			
Northeast	18	18	16
Midwest	22	22	21
South	36	36	39
West	25	24	24
Parent's education level			
Grade 11 or less	14	13	11
Grade 12 or GED	29	29	19
1-3 years of college	31	30	26
4+ years of college	27	27	44
Educational status			
General education	92	91	87
Special education	8	9	13

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

^aExcept where noted, estimates derived from the 2016–2017 digest of educational statistics (Snyder et al., 2019). ^bDerived from the most recent estimates from the 2013–2014 civil rights data collection survey (U.S. Department of Education Office for Civil Rights, 2016).

scales. Brief descriptions of these rating scales as well as the SEL*b* and EBC scoring frameworks follow.

SSIS Problem Behavior Scales

The SSIS-RS (Gresham & Elliott, 2007) is used by professionals to screen and classify the social behavior of students 3–18 years of age. The SSIS-RS features a multirater (parents, teachers, and students with at least 3rd grade reading ability) approach that provides a comprehensive examination of seven areas of prosocial skills (communication, cooperation, assertion, responsibility, empathy, engagement, and self-control) and five areas of problem behaviors (internalizing, externalizing, bullying, hyperactivity/inattention, and autism spectrum). The instrument yields norm-referenced scores based on a national sample representative of the 2006 US Census. The SSIS-RS Manual provides extensive validity evidence based on test content, internal structure, inter-correlations among scales and subscales, item-total correlations, and relations with other variables (Gresham & Elliott, 2008). Item development was based on a broad review of the empirical literature on social skills difficulties in special populations, reviews of published empirical studies using an earlier version of the scale (Social Skills Rating System, Gresham & Elliott,

1990), and research on the relationship between specific social behaviors and important social outcomes for children and youth.

For the current study, the Problem Behavior items of the SSIS-RS served as the initial item pool for the development of the Emotional Behavior Concerns (EBC) scales. Specifically, items were selected for the EBC Scales from the SSIS-RS Internalizing and Externalizing items utilizing the approach described in the analyses section. Furthermore, due to the conceptual link between bullying and externalizing behavior and the practical importance of identifying and addressing bullying, we added all SSIS-RS Bullying items into the Externalizing item pool for our analyses (i.e., the targeted domain encompassed both Externalizing behavior and Bullying).

In addition to selecting the items for the EBC scales, we developed a criterion-referenced concern framework (CRCF) to facilitate interpretation of the raw scores from the EBC Internalizing and the Externalizing Scales. A three-level *No Concern*, *Possible Concern*, and *Concern* model was operationalized with cut-scores influenced by mean score distributions for the students representing different racial/ethnic and gender subgroups from the large SSIS-RS standardization sample (See Figure 1). For

Figure 1. The CRCF Descriptions for the EBC-I and EBC-E Scores. Reprinted With Permission

Emotional Behavior Concern – Internalizing (EBC-I)	Emotional Behavior Concern – Externalizing (EBC-E)
<p>No Concern Level Students at this level rarely (never or seldom) ...</p> <ul style="list-style-type: none"> • Comment about negative events in life, • Act worried, • Feel down, • Act lonely, • State others don't like them. <p>These students do not require additional attention or support at this time.</p>	<p>No Concern Level Students at this level rarely (never to seldom) ...</p> <ul style="list-style-type: none"> • Persuade others to do something they dislike doing, • Verbally argue with others, • Act without thinking things through, • Break agreed-upon rules, • Keep others out of social groups. <p>These students do not require additional attention or support at this time.</p>
<p>Possible Concern Level Students at this level sometimes (seldom or often) ...</p> <ul style="list-style-type: none"> • Comment about bad things happening, • Act nervous interacting with others, • Feel sad, • Act lonely, • State others don't care about them. <p>These students may need additional attention and some focused support, along with periodic monitoring of their potential internalizing concern(s) to ensure the situation does not worsen.</p>	<p>Possible Concern Level Students at this level sometimes (seldom to often) ...</p> <ul style="list-style-type: none"> • Pressure others to do things against their will, • Fight physically with others, • Act impulsively, • Break rules intentionally, • Exclude others from social groups. <p>These students may need additional attention and some focused support, along with periodic monitoring of their potential externalizing concern(s) to ensure the situation does not worsen.</p>
<p>Concern Level Students at this level frequently (often or almost always) ...</p> <ul style="list-style-type: none"> • Talk about bad things happening to them, • Act anxious interacting with others, • Feel depressed, • Act lonely, • State few people care about them. <p>These students likely need school professionals' ongoing support and monitoring, until a comprehensive assessment of their internalizing concern(s) can be completed and appropriate support services determined.</p>	<p>Concern Level Students at this level frequently (often to almost always) ...</p> <ul style="list-style-type: none"> • Force others to do things against their will, • Fight physically with others, • Act impulsively, • Break rules intentionally, • Isolate others from groups. <p>These students likely need ongoing support and monitoring, until a comprehensive assessment of their externalizing concern(s) can be completed and appropriate support services determined.</p>

future reference on the development of this framework, see Appendix D of the User Guide & Technical Manual (Elliott et al., 2020b).

SSIS SEL Brief (SSIS SELb) Scales

The Teacher and Student K-12 versions of the SSIS SELb Scales (Elliott et al., 2020a) both consist of 20 items (17 of which were originally used on the SSIS-RS) with four-item scales measuring Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision Making. The Teacher version is for students in Grades K-12, while the Student version with a readability level of 2.5 is appropriated for students in Grades 3–12. Both these measures are administered online and require less than 5 min to complete.

A brief summary of the reliability (internal consistency, test-retest, inter-rater, and IRT test information or TIF estimates) and validity (internal structure, and relationship with related constructs) evidence for the SSIS SELb-T scores (e.g., Anthony et al., 2020a, 2020b) indicates they are sufficiently precise for low stakes decisions (TIF ≥ 5 which is roughly equivalent to reliability $\geq .80$) from roughly -3 to roughly 0.75 on the respective latent trait standard scale. In addition, for the Composite score, the Cronbach's α was $.95$, the test-retest reliability coefficient was $.78$, and the interrater reliability coefficient for pairs of teachers rating the same student was $.65$. With regard to internal structure validity evidence, SSIS SELb-T scale scores were moderately to strongly inter-related. For evidence regarding relationship with related constructs, SSIS SELb-T scores were moderately to strongly negatively correlated with BASC-2 scales that reflect emotional or behavioral difficulties as expected. SSIS SELb-T scores also were moderately to strongly positively related to BASC-2 scales that assess adaptive skills, communication, socialization, and general adaptive behavior.

Similarly, a summary of the reliability (internal consistency, test-retest, inter-rater, and TIF estimates) and validity (internal structure, and relationship with related constructs) evidence for the SSIS SELb-S Composite score indicates Cronbach's α was $.90$ and test-retest reliability was $.87$. IRT based information indicated the SSIS SELb-S scales maintained a $.70$ level of reliability across broad levels of each SEL construct. With regard to structural and relational validity evidence, the SSIS SELb-S inter-scale correlations were all moderate ($.55 - .65$) and correlations with the BASC-2 followed expected patterns with inter-scale correlations ranging from $.52$ to $.84$ (median = $.61$) across the SSIS SELb-S Composite and SSIS SELb-S scales.

A competency-based criterion-reference performance framework (CRPF) consistent with the five CASEL

competency domains is used to interpret the Composite Score from each SELb assessment. Specifically, this strength-focused, competency-referenced approach characterizes clusters of self-awareness, self-management, social awareness, relationship, and responsible decision-making skills into four performance levels: *Emerging*, *Developing*, *Competent*, and *Advanced*. The Competent level of performance of the SEL CRPF is presented as an example in Figure 2. Each performance level is intended to vary developmentally in comparison to the next higher level. Thus, across the performance levels from Emerging to Advanced, there is a progression of fundamental SEL skills that occur more frequently, with less support, and in more social situations or environments. The same general progressions of skills are expected for each level and collectively contribute to one's composite SEL performance level. A detailed account of the development of the SEL CRPF is provided in the User Guide & Technical Manual of the SSIS SEL Brief Scales (Elliott et al., 2020a) and in a Technical Report on the Development and Verification of the SEL CRPF posted at SSIScolab.com.

Behavior Assessment System for Children-Second Edition

The BASC-2 (Reynolds & Kamphaus, 2004) was used as a concurrent validity measure with a subsample of students ($n=57$). The BASC-2 Teacher Rating Scale is a

Figure 2. The CRPF Competence Level Description

Competent Level
Students at this performance level often exhibit most of the following competencies appropriately and with minimal or no prompting in social situations:
<ul style="list-style-type: none"> • Accurately recognize one's emotions, 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. • 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. • Take the perspective of and empathize with others from diverse backgrounds and cultures, understand social and ethical norms for behavior, and recognize family, school, and community resources and supports. • 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. • Make constructive and respectful choices about personal behavior and social interactions based on consideration of safety concerns, social norms, the realistic evaluation of consequences of various actions, and the well-being of self and others.

Reprinted with Permission. Source: SSIS SEL Brief + Mental Health Scales User Guide & Technical Manual (Elliott et al., 2020c)

norm-referenced assessment including 16 primary scales and five composite scales (Adaptive Skills, Behavioral Symptoms Index, Externalizing Problems, Internalizing Problems, and School Problems). The BASC-2 manual reports internal consistency coefficients in the .90s for the composite scales for both a general sample and a clinical sample. Retest (1–8 weeks) reliability with the BASC-2 yielded average correlations in the .80s for composite scores. Substantial validity evidence for BASC-2 scores in the form of convergent and discriminant validity correlations is reported in the assessment's Technical Manual (Reynolds & Kamphaus, 2004).

Procedures

The data used to develop the EBC scales were collected as part of the original SSIS-RS standardization. Field staff from Pearson Assessment recruited school site coordinators in 115 schools across 36 states. Site coordinators managed data collection from fall 2006 to fall 2007. No missing data were reported. The final original sample was selected from the larger respondent sample to fit the 2006 U.S. Census demographics of age, sex, race/ethnicity, and educational status.

Data Analyses

The development of the EBC scales from the SSIS-RS Problem Behavior items relied on IRT methods and followed similar procedures to those used for the development of the SSIS SEL Brief Scales (Anthony et al., 2020a, 2020b, 2020c). Briefly, assumptions for IRT (unidimensionality and local independence) were assessed for each of the SSIS-RS Internalizing and Externalizing scales using ordinal exploratory factor analysis (EFA) in Mplus (Muthén & Muthén, 1998–2019) and chi-squared statistics of local dependence (LD χ^2) from fitting the Graded Response Model (GRM; Samejima, 1969) in IRTpro (Cai et al., 2019). Essential unidimensionality was met if ratios of the first to the second eigenvalues from the EFA > 4 (Reeve et al., 2007) and RMSEA for the single-factor model $< .10$ (MacCallum et al., 1996). Low-loading items were removed until these criteria were met. Local independence was considered violated if standardized LD $\chi^2 > 10$ (Cai et al., 2019), and violation was addressed during item selection by not including both in the final scales.

Moreover, differential item functioning (DIF) across sex (girls vs. boys) and race/ethnicity (white vs. nonwhite) was assessed using both statistical tests (including a two-step purification process; Tay et al., 2014) and effect sizes to detect item bias. We used the Expected Score

Standardized Difference (Meade, 2010) as a measure of effect size, which follows Cohen's criteria of .2, .5, and .8 for small, medium, and large effects, respectively. Items were selected to balance five criteria: content representation of the construct, their contribution to scale information (analogous to reliability), lack of sex-DIF, lack of race-ethnicity DIF, and lack of LD with other items on the scale. Based on this goal, each of the authors identified 5–6 items independently, and disagreements were resolved by group discussion to finalize the selection of 5 items each for the EBC-I and EBC-E scale.

Initial Reliability Analyses

Several reliability indexes were examined for each of the EBC scales, including test information function (TIF), Cronbach's α , omega hierarchical (McDonald, 1999) based on the hypothesized two-factor structure (see validity analyses below), test-retest reliability coefficients, and interrater (between teachers for the Teacher form) reliability coefficients. Samples for the Teacher form were 750, 144, and 54 students for α/ω , test-retest, and interrater coefficients, respectively. For the Student form, sample size was 800 for α/ω and 127 for test-retest coefficient.

Initial Validity Analyses

First, a standard two-factor confirmatory factor analysis (CFA) model where items scores were treated as categorical with the WLSMV estimator in Mplus was used to verify the factor structure of the EBC scales (i.e., EBC-I items loading on one factor and EBC-E items loading on the other). Intercorrelation between EBC-I and EBC-E scales were then examined, followed by correlations between EBC scales and all SSIS SELb scales as well as the composite scale. As the EBC scales assess negative behaviors and SELb scales assess positive ones, negative correlations were expected between these scales. Correlations also were calculated between ratings on EBC scales and BASC-2 scale scores with moderate positive relationships expected between EBC scale scores and negative-valence BASC-2 scales (e.g., School Problems) and moderate negative relationships between EBC scale scores and positive-valence BASC-2 scales (e.g., Adaptive Skills). Finally, the utility of applying both the SEL CRPF and EBC CRCF was examined to determine whether information gleaned from each set of scales were unique. Specifically, the percentages of students falling within each EBC CRCF risk category by SEL CRPF competency category were tabulated and EBC score distributions were plotted by SEL competency levels for the full standardization sample.

RESULTS

Item Selection

Essential unidimensionality was adequately satisfied for the Student EBC-I item pool (ratio of 1st to 2nd eigenvalues = 6.07, RMSEA = .05) and EBC-E item pool (ratio of 1st to 2nd eigenvalues = 7.72, RMSEA = .03). The Teacher EBC-I item pool (ratio of 1st to 2nd eigenvalues = 6.77, RMSEA = .03) and EBC-E item pool (ratio of 1st to 2nd eigenvalues = 14.78, RMSEA = .03) also demonstrated essential unidimensionality.

Using the selection process and criteria described previously, five items were identified to represent the internalizing behavior construct for the EBC-I scale. These included items involving withdrawing from others; acting lonely, sad, or anxious; and making negative self-statements. Similarly, five items were also selected to represent the externalizing behavior construct for the EBC-E scale. These included items involving acting impulsive, fighting with others, disobeying rules, forcing others to act against their will, and excluding others from social groups.

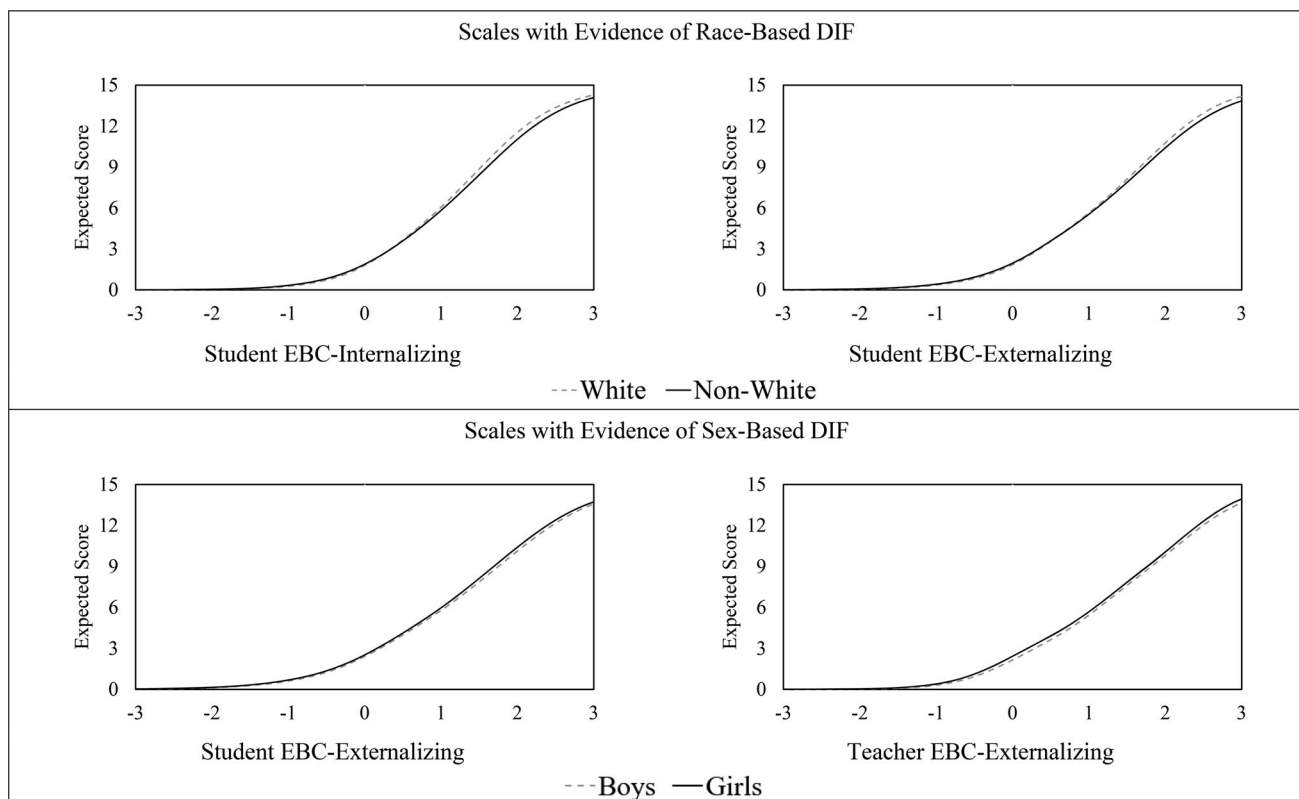
With respect to DIF, Student EBC-I items were relatively free of sex- and race-DIF except for one item “I feel lonely” that showed negligible race-DIF ($ES = -.07$ with slightly higher expected scores for White). The Student EBC-E scale contained one item “I often do things without

thinking” with small sex-DIF ($ES = .21$ with higher expected scores for girls) and one item “I make people do what I want them to do” with negligible race-DIF ($ES = .02$ with slightly higher expected scores for non-White). The Teacher EBC-E scale had no race-DIF items, but one item “keeps others out of social circles” showed small-medium sex-DIF ($ES = .40$ with higher expected scores for girls). Given the small number of DIF items and relatively small effect sizes for the few that showed statistical DIF (likely due to large sample sizes), the effect of including them in the EBC scales was negligible (as shown in Figure 3). Moreover, none of the items included in EBC-I and EBC-E, Teacher or Student form, showed LD issues.

Reliability Analyses

Score reliability estimates for the EBC-I and EBC-E Scales were evaluated using internal consistency (Cronbach's α , and omega hierarchical based on the two-factor model), test-retest, and interrater reliability coefficients (Table 2). As documented, internal consistency (α and omega hierarchical), and test-retest values for both these scales on the Teacher form were at or exceeded the .80 criterion commonly used for screening decisions (Salvia et al., 2016). These same reliability estimates were slightly lower for the Student Form. In addition, teacher-teacher interrater

Figure 3. Expected Scores Across Focal and Reference Groups for Retained EBC Scales with Statistical Evidence of DIF



reliability coefficients were moderate, which is consistent with inter-informant agreement research (e.g., Gresham, Elliott, Metallo, Byrd, Wilson, & Cassidy, 2018; Rupp et al., 2018) for social behavior assessments.

As shown in Figure 4, the Test Information Functions (TIFs) for each Teacher and Student EBC Scale indicate

that score precision met or exceeded the criterion of .8 across a broad spectrum of internalizing and externalizing concerns (approximately above $-.5$ on the standard latent trait scale). These results show a high level of precision at the upper ranges of EBC-I and EBC-E scores typical of students in need of additional support.

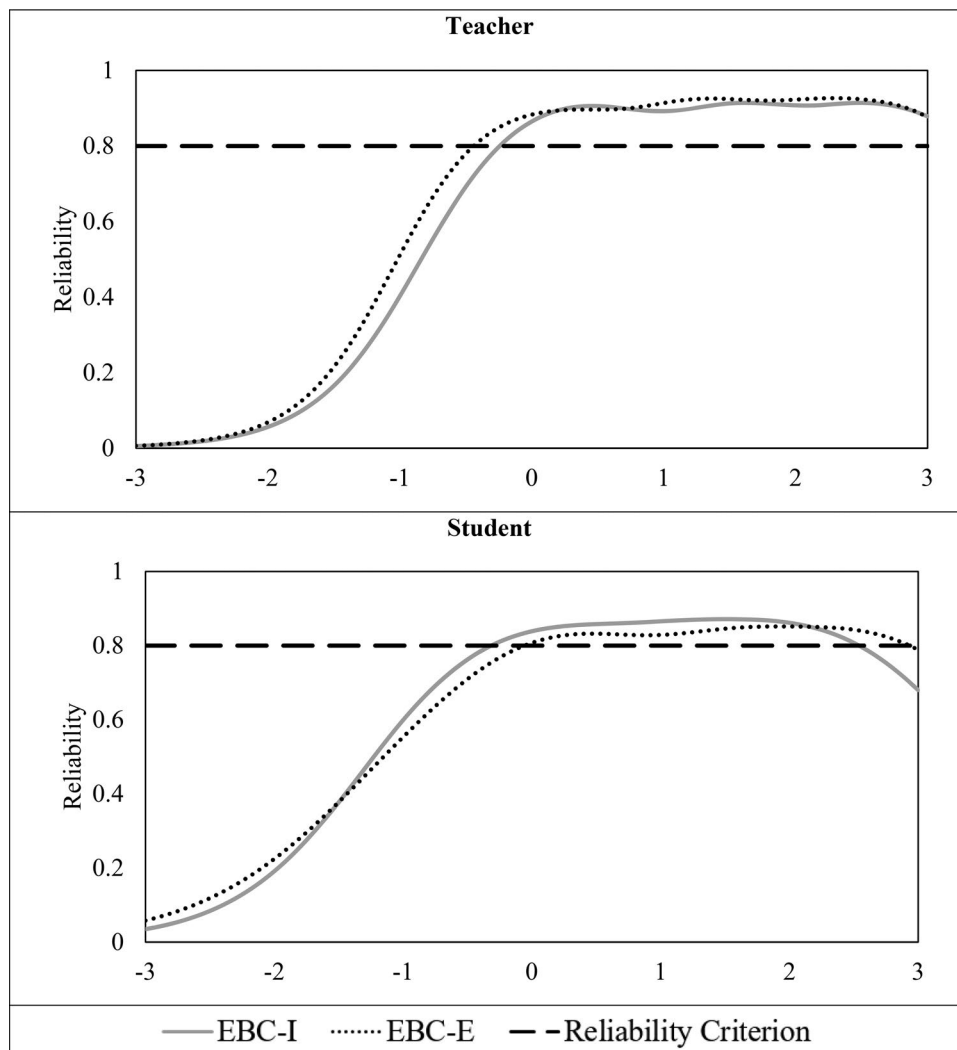
Table 2. Reliability Statistics for EBC Scales Across Informants

Informant	Cronbach's α (Omega H)		Test-retest Reliability		Interrater Reliability	
	EBC-I	EBC-E	EBC-I	EBC-E	EBC-I	EBC-E
Teacher	.84 (.85) ($n = 750$)	.85 (.86) ($n = 750$)	.83 ($n = 144$)	.80 ($n = 144$)	.46 ($n = 54$)	.39* ($n = 54$)
Student	.79 (.80) ($n = 800$)	.75 (.76) ($n = 800$)	.60 ($n = 127$)	.64 ($n = 127$)	—	—

Note. Omega H = Omega hierarchical (McDonald, 1999) based on a standard two-factor CFA model; EBC-I = Emotional Behavior Concern—Internalizing; EBC-E = Emotional Behavior Concern—Externalizing. Unless otherwise noted all correlations statistically significant ($p < .001$).

* $p < .01$.

Figure 4. Test Information Functions for EBC-I and EBC-E Scales Across Informant



Note. EBC-I = Emotional Behavior Concern—Internalizing; EBC-E = Emotional Behavior Concern—Externalizing. Test Information Converted to Reliability Metric through Formula from Petrillo et al. (2015). Reprinted with permission from SSIS SEL Brief + Mental Health Scales User Guide & Technical Manual

Validity Analyses

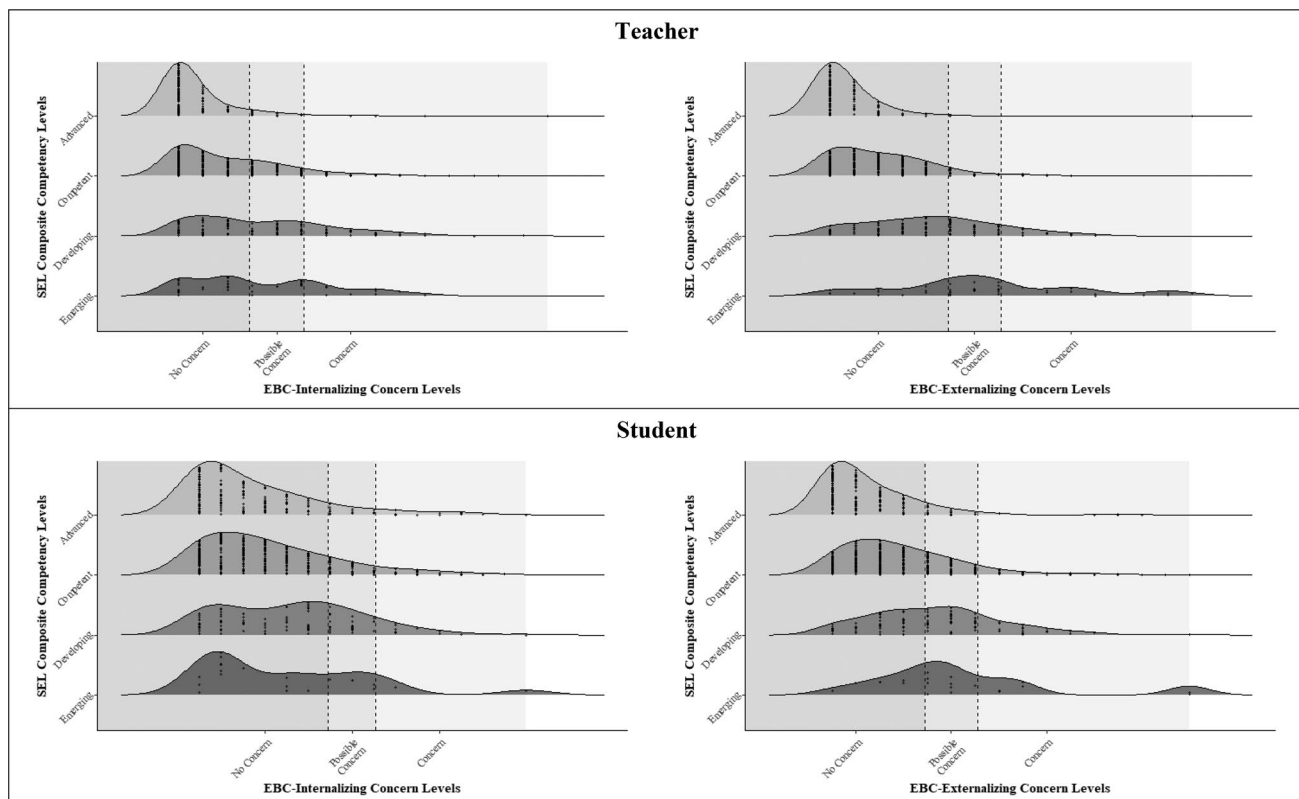
Several forms of validity evidence were examined including internal structure (CFA, intercorrelations between EBC-I and EBC-E Scales), and relationship with related constructs (correlations between the EBC Scales and SSIS SELb Scales, and correlations between the EBC Scales and the BASC-2). First, the hypothesized standard two-factor model with EBC-I items loading on one factor and EBC-E items loading on the other fit both the student data ($\chi^2(34)=121.169$, CFI=.980, RMSEA=.057, SRMR=.035) and the teacher data ($\chi^2(34)=236.205$, CFI=.971, RMSEA=.089, SRMR=.053) reasonably well. All standardized factor loadings were high for the teacher sample ($\geq .73$) and slightly lower for the student sample ($\geq .67$). Estimated intercorrelations between the EBC-I and EBC-E factors were .59 for teacher and .71 for student. The correlations between EBC-I and EBC-E observed scores as rated by teachers was .45 and as rated by students .54; both positive and indicative of moderate overlap between the constructs as perceived especially by students.

Second, the intercorrelations between the EBC scales, the SELb Composite score, and five SELb scales were as expected. That is, all the correlations were negative between SELb scales and both the EBC-I (e.g., $-.36$ Teacher EBC-I and SEL Composite; $-.18$ Student EBC-I and SEL

Composite) and EBC-E (e.g., $-.60$ Teacher EBC-E and SEL Composite; $-.40$ Student EBC-E and SEL Composite). The magnitude was moderate to high between the EBC-E scale and SEL scales ($-.24$ to $-.70$ across Teacher and Student forms), while being low to moderate between the EBC-I scale and SEL scales ($-.08$ to $-.40$ across Teacher and Student forms). This pattern generally was consistent for both Teacher and Student forms.

Third, additional relational validity evidence was examined in the form of correlations between the EBC scales and the BASC-2. Although the sample of teachers and students who completed ratings on both these assessments was relatively small, the pattern and magnitude of the correlations were as expected. That is, the EBC-I and EBC-E scales correlated positively and moderately to highly (Teacher correlation range .33 to .88; Student correlation range .26 to .69) with all the BASC-2 scales except the Adaptive Skills/Personal Adjustment scale which was correlated negatively and moderately (Teacher EBC-I $-.67$ and EBC-E $-.57$; Student EBC-I $-.56$ and EBC-E $-.37$) given this scale measures desired behaviors. Also of note, the EBC-I as rated by both teachers and students correlated strongly with the BASC-2 Internationalizing scale and the Behavior Symptoms/Emotional Symptoms scale (Teacher .71 and .81, respectively; Student .64 and .69,

Figure 5. EBC-Internalizing and EBC-Externalizing Score Distributions by SEL Composite Levels for Teacher and Student Forms



respectively), and the EBC-E correlated highest with the BASC-2 Externalizing/Inattention-Hyperactivity scale (Teacher .88; Student .51).

The Utility of Concurrent SEL and EBC Assessment

Finally, we used the SSIS SEL*b* and EBC scoring frameworks to evaluate the utility of concurrently assessing both SEL skills and EBCs. A diagram of the breakdown of EBC score distributions by SSIS SEL*b* SEL Competency Levels is found in Figure 5, and percentages of students falling in combinations of SEL competency levels and EBC concern levels can be found in Table 3 (with highlighting to focus attention on key outcomes). As can be seen, there were a number of students rated by teachers in the SEL Competent ranges who showed concurrent EBC-I (4.5%) and EBC-E (1.3%) difficulties at the Concern Level. Similarly, a number of students rated themselves in the Competent range *and* the Concern range for EBC-I (4.1%) and EBC-E (4.6%).

DISCUSSION

Children exhibit a range of social emotional behaviors, yet virtually all SEL assessments and intervention programs focus on the positive or strength-focused subset of children's emotions and behaviors. This study examined the reliability and validity evidence for two new scales intended to efficiently assess students' emotional behavior concerns along with the SEL skills assessed by the SSIS SEL*b* Teacher and Student Forms. In addition, it explored the utility of concurrent SEL and EBC assessment using criterion-referenced scoring frameworks for interpreting EBC alongside the SEL competencies of K-12 students. As a result, this study extended a program of research on the efficient assessment of students' social emotional skills and behaviors by collecting evidence from teachers and students themselves concerning cooccurring EBC.

Key Findings

Reliability and Validity Evidence of the EBC Scales

Overall, initial evidence was positive for the EBC-I and EBC-E Scales across teacher and student informants. Estimates of reliability (that included internal consistency, test-retest, inter-rater, and TIF estimates) indicated sufficient precision for low-stakes decision making. Furthermore, as indicated by scale TIFs, score reliability for the EBC scales was highest precisely in score ranges most likely to indicate the presence of behavior concerns that need to be further assessed. With regard to concurrent relation validity evidence, all correlations with the scales on the SSIS SEL*b* and those of the BASC-2 were consistent with expectations and supported the validity of scores from the EBC Scales. For example, correlations between the EBC-E and BASC-2 Externalizing/Hyperactivity-Inattention scales were higher than the EBC-I correlations with these scales. The same pattern held for the EBC-I—BASC-2 Internalizing scale correlations. Such findings offer strong initial evidence of validity of scores from the EBC-I and EBC-E Scales for universal screening.

Utility of the Criterion-Reference Scoring Frameworks

The CASEL Competency Framework was the cornerstone for the development of the SEL*b* and its competency-referenced performance framework to characterize students' SEL developmental status. Similarly, a criterion-referenced concerns framework was developed to characterize levels of concern associated with students' internalizing and externalizing EBC scores. Together, the raw SEL Composite, EBC-I, and EBC-E scores were meaningfully transformed via these scoring frameworks to provide a broader picture of the social emotional functioning of students. Clearly, the EBC scales provide functional information beyond that from the strength focused SEL competency performance information alone. The

Table 3. Numbers of Students (and Percentages of Total Sample) by EBC Concern Level, SEL Competency Level, and Informant

SEL competency level	EBC-I concern levels			EBC-E concern levels		
	No concern	Possible concern	Concern	No concern	Possible concern	Concern
Teacher						
Emerging	23 (3.1)	9 (1.2)	9 (1.2)	17 (2.3)	12 (1.6)	12 (1.6)
Developing	74 (9.9)	33 (4.4)	31 (4.1)	96 (12.8)	25 (3.3)	17 (2.3)
Competent	300 (40.0)	60 (8.0)	34 (4.5)	371 (49.5)	13 (1.7)	10 (1.3)
Advanced	166 (22.1)	7 (0.9)	4 (0.5)	176 (23.5)	0 (0.0)	1 (0.1)
Student						
Emerging	15 (1.9)	4 (0.5)	2 (0.3)	10 (1.3)	5 (0.6)	6 (0.8)
Developing	80 (10.0)	15 (1.9)	11 (1.4)	55 (6.9)	31 (3.9)	20 (2.5)
Competent	372 (46.5)	30 (3.8)	33 (4.1)	339 (42.4)	59 (7.4)	37 (4.6)
Advanced	213 (26.6)	9 (1.1)	16 (2.0)	219 (27.4)	12 (1.5)	7 (0.9)

Note. Percentages in parentheses. Not all percentages within each quadrant of the table sum exactly to 100 due to rounding. Shading indicates when discrepancies exist across EBC and SEL levels with the darkest shade indicating where levels are most disparate.

additional information regarding EBC provides more insight regarding risk than only knowing students' SEL competence level status and vice versa. For example, across both raters, 3.1% to 14.4% of students scored in the Competent or Advanced SEL range and the Possible Concern or Concern level of the EBC scales. This clearly indicates that the assumption that EBCs are "covered" by SEL-focused assessments does not hold, at least with the SSIS SELb. Furthermore, being rated in the EBC Possible Concerns or Concerns range also did not guarantee limited SEL competencies. For example, of children rated in the Possible Concern or Concern range on EBC scales, 4.1% to 10.9% were also rated in the Emerging or Developing SEL range. Such findings clearly indicate the importance of concurrent assessment of both SEL skills and EBC.

Limitations and Future Research

There are several limitations to the current study that need to be addressed with future research. First, the prospective item pool for the EBC scales was constrained to the items included in the original SSIS-RS Problem Behavior Scales. Although multiple criteria were used to inform the selection of an efficient set of items that also reflected the breadth of the target constructs (internalizing & externalizing/bullying), testing additional items beyond those included in the original Problem Behavior scales will help ensure that the items on the EBCs appropriately represent their intended constructs. Relatedly, while the size of the original SSIS standardization sample was large enough to construct samples for the current study that still reflected the current diversity of the US student population, new samples will provide additional evidence to evaluate the generalization of psychometric properties of scores from the EBC scales. In addition, the inclusion of larger samples and additional measures of convergent constructs (e.g., depression & anxiety for the EBC-I scale) will provide further insight regarding functioning of the scale and its scores.

Beyond addressing these limitations, there are three other important directions for future research related to the use of the EBC scales as well as other rating scales focused on challenging behaviors or featuring multiple informants. The first is examining the implications of decision-rules based on score ranges like those within the criterion-referenced scoring frameworks of the SELb and EBC scales. Though the development and application of such frameworks have become standard practice in large scale assessments of English language arts and mathematics, they are far less common in the SEL domain and with indirect forms of assessment such as behavior rating scales.

As such, an important area of research is examining how best to develop interpretive systems in the SEL domain that are accurate, fair, and provide actionable data for students and teachers.

A related direction for future research regarding universal screening assessments is examining potential unintended consequences of identifying students' skills or behaviors as falling within a "possible concern" level or range. Incorporating such a level within the EBC criterion-referenced framework was intended to *prevent* students from being identified as having problems when scores are not definitively indicative of such status after accounting for measurement error. Instead, additional assessment and monitoring may be necessary to determine if a problem truly exists or emerges over time. If users are unable to implement such an approach in schools due to time, other resource constraints, etc., it could result in students being either (a) identified as having a problem when there is none or (b) inappropriately denied supports when they actually need them.

One method to potentially minimize the likelihood of such unintended consequences is to incorporate multiple data sources into the decision-making process. The EBC and SELb scales include multiple forms that facilitate inclusion of multiple perspectives. As shown in the current study (and indicated in previous studies of cross-informant agreement), multiinformant data tend to demonstrate small-to-moderate relationships. As such, challenges can arise when attempting to draw conclusions about plans of action based on conflicting multiinformant data. Future research related to the EBC and similar scales is necessary to determine how best to interpret/integrate multiinformant data to yield the best decisions and outcomes for students.

Implications for Practice

The creation and initial validation of the EBC scales have direct implications for practice. Given growing recognition and concern about the interface between student mental health, learning, and academic outcomes, there is need for broad, yet efficient, assessments that can be used to screen the whole social emotional child. Although brief scales, such as the DESSA-mini, have been developed to quickly assess students' strength-focused social emotional functioning, they do not consider the possibility of co-occurring emotional or behavioral concerns. Evidence from the current study demonstrates that such an omission could result in 3% to 15% of the student population being considered as socially emotionally "healthy" while experiencing an emotional or behavioral concern. Conversely, if teams were to solely screen for emotional behavior

concerns, data from the current study suggest that they would miss the fact that most students with concerns have social emotional skills that can be drawn upon from an intervention planning standpoint. To efficiently gather insights in both domains, practitioners can use the SSIS SELb + Mental Health Scales (SSIS SELb+MHS; Elliott et al., 2020b), which represents the concurrent use of the SSIS SELb scales and the EBC scales examined in this study. Given their relative brevity, time should not be a barrier to the use of the expanded mental health focus of the SEL Brief Scales. And given that SEL skills are the primary focus of the scales, concerns about mental health screening may be minimized.

CONCLUSIONS

This study is premised on the belief that SEL competencies are important in the lives of children and that the CASEL Competency Framework has positively influenced the advancement of SEL-focused programs in many schools. An essential part of these programs is the universal screening of students' SEL skills indicative of core competencies with the goal of identifying strengths and areas in need of improvement. Advancing children's wellbeing, however, involves more than improving SEL skills; for some children it also means addressing emotional behavior concerns, which are not simply synonymous with low levels of SEL skills. These emotional behavior concerns also can be efficiently screened for large groups of students. Thus, it makes sense to consider the use of an assessment that efficiently integrates screening of SEL and EBC.

The current study considered just such an integrated screener and was designed to answer two questions "Is there substantial evidence that the scores from the EBC Scales of the SSIS SEL Brief + Mental Health Scales are reliable and valid for universal screening?" and more importantly "Does the assessment of students' emotional behavior concerns add information or simply duplicate information provided by SEL scores alone?" With regard to the first research question, results suggest that scores from the EBC-I and EBC-E Scales demonstrate adequate reliability (as estimated by internal consistency, test-retest, inter-rater correlations, and TIF), for initial, low-stakes screening decisions. In addition, EBC scores relate as expected with established comprehensive measures of the same or highly similar constructs (i.e., BASC-2 scores). They also relate as expected with scores from positively-framed (and CASEL framework-aligned) measures of positive social-emotional skills and demonstrate low-to-moderate agreement across informants.

With regard to the second question, results from the current study suggest that, though important to

promoting student mental health, focusing exclusively on positive social-emotional skills within a universal screening system likely would miss a number of students who are experiencing significant emotional-behavioral concerns and potentially in need of intervention. The EBC scales address this limitation by providing brief multiinformant measures to efficiently assess emotional behavior concerns at the universal level. This combined approach with items based on a common standardization sample provides important insights that can be used to advance the social emotional health of the whole social emotional child. As such the SSIS SELb+MHS (Elliott et al., 2020c), which includes the brief SEL and EBC Scales, was created based on the results of this research to advance more efficient and effective socioemotional service delivery in schools.

DISCLOSURE

The authors of this article are also the authors of the SSIS SEL Brief + Mental Health Scales published by SAIL CoLab. They receive a royalty for the sale of these rating scales.

ORCID

Pui-Wa Lei  <https://orcid.org/0000-0002-8179-6623>

REFERENCES

- Anthony, C. J., & DiPerna, J. C. (2017). Identifying sets of maximally efficient items from the Academic Competence Evaluation Scales-Teacher Form. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, 32(4), 552–559. <https://doi.org/10.1037/spq0000205>
- Anthony, C. J., & DiPerna, J. C. (2018). Piloting a short form of the academic competence evaluation scales. *School Mental Health*, 10(3), 314–321. <https://doi.org/10.1007/s12310-018-9254-7>
- 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. <https://doi.org/10.1177/0734282917743335>
- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P.-W. (2020a). The SSIS SEL brief scales-student form: Initial development and validation. *School Psychology (Washington, DC)*, 35(4), 277–283. <https://doi.org/10.1037/spq0000390>
- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P.-W. (2020b). Multi-rater assessment of young children's social and emotional learning via the SSIS SEL brief scales - preschool forms. *Early Childhood Research Quarterly*, 53, 625–637. <https://doi.org/10.1016/j.ecresq.2020.07.006>
- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P.-W. (2020c). Initial development and validation of the SSIS SEL brief

- scales - teacher form. *Journal of Psychoeducational Assessment*.
- Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K-12 schools. *Education and Treatment of Children*, 37(4), 611–634. <https://doi.org/10.1353/etc.2014.0039>
- Cai, L., Thissen, D., & Du Toit, S. H. C. (2019). *IRTPRO for Windows [Computer software]*. Scientific Software International.
- CASEL. (2003). Safe and sound: An educational leader's guide to evidence-based social and emotional learning (SEL) programs. Chicago: CASEL. <https://casel.org/wp-content/uploads/2016/01/PDF-16-safe-and-sound.pdf>
- CASEL. (2020). *CASEL guide: PreK and elementary evidence-based programs*. Retrieved July 25, 2020, from www.casel.org/guide/programs
- Crowder, M. K., Gordon, R. A., Brown, R. D., Davidson, L. A., & Domitrovich, C. E. (2019). Linking social and emotional learning standards to the WCSSD social-emotional competency assessment: A Rasch approach. *School Psychology (Washington, DC)*, 34(3), 281–295. <https://doi.org/10.1037/spq0000308>
- Denham, S. A. (2015). Assessment of SEL in educational contexts. In J. Durlak, C. Domitrovich, R. Weissberg, & T. Gullotta (Eds.), *Handbook of social & emotional learning: Research & practice* (pp. 285–300). Guilford.
- DiPerna, J. C., Lei, P., Bellinger, J., & Cheng, W. (2015). Efficacy of the social skills improvement system classwide intervention program (SSIS-CIP) primary version. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, 30(1), 123–141. <https://doi.org/10.1037/spq0000079>
- DiPerna, J. C., Lei, P., Bellinger, J., & Cheng, W. (2016). Effects of a universal positive classroom behavior program on student learning. *Psychology in the Schools*, 53(2), 189–203. <https://doi.org/10.1002/pits.21891>
- DiPerna, J. C., Lei, P., Cheng, W., Hart, S. C., & Bellinger, J. (2018). A cluster randomized trial of the social skills improvement system-classwide intervention program (SSIS-CIP) in first grade. *Journal of Educational Psychology*, 110(1), 1–16. <https://doi.org/10.1037/edu0000191>
- DiPerna, J. C., Volpe, R., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review*, 31(3), 298–312. <https://doi.org/10.1080/02796015.2002.12086157>
- DiPerna, J. C., Volpe, R., & Elliott, S. N. (2005). A model of academic enablers and math achievement in elementary grades. *Journal of School Psychology*, 43(5), 379–392. <https://doi.org/10.1016/j.jsp.2005.09.002>
- Dowdy, E., Doane, K., Eklund, K., & Dever, B. (2013). A comparison of teacher nomination and screening to identify behavioral and emotional risk within a sample of underrepresented students. *Journal of Emotional and Behavioral Disorders*, 21(2), 127–137. <https://doi.org/10.1177/1063426611417627>
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Dusenbury, L., Yoder, N., Dermody, C., & Weissberg, R. (2019). An examination of frameworks for social and emotional learning (SEL) reflected in state K-12 learning standards. *Measuring SEL: Using Data to Inspire Practice*. CASEL.
- Elliott, S. N., Anthony, C. J., DiPerna, J. C., & Lei, P. W. (2020). *The SEL Competency-Referenced Performance Framework*. Scottsdale, AZ: SAIL Collaborative.
- Elliott, S. N., Anthony, C. J., DiPerna, J. C., Lei, P. W., & Gresham, F. M. (2020b). *SSIS SEL brief Scales - Preschool Parent*. Scottsdale, AZ: SAIL Collaborative.
- Elliott, S. N., Anthony, C. J., DiPerna, J. C., Lei, P. W., & Gresham, F. M. (2020c). *SSIS SEL Brief + Mental Health Scales - Teacher*. Scottsdale, AZ: SAIL Collaborative.
- Elliott, S. N., & Gresham, F. M. (2008). *SSIS intervention guide*. Pearson.
- Elliott, S. N., & Gresham, F. M. (2020). *SSIS SEL classwide intervention program (CIP)*. SAIL CoLab.
- Forness, S. R., Kim, J., & Walker, H. M. (2012). Prevalence of students with EBD: Impact on general education. *Beyond Behavior*, 21, 3–9.
- Greenspoon, P. J., & Saklofske, D. H. (2001). Toward an integration of subjective well-being and psychopathology. *Social Indicators Research*, 54(1), 81–108. <https://doi.org/10.1023/A:1007219227883>
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating scales*. Pearson.
- Gresham, F. M., & Elliott, S. N. (2007). *Social Skills Improvement System: Rating Scales*. Bloomington, MN: Pearson Assessments.
- Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales*. Pearson.
- Gresham, F. M., & Elliott, S. N. (2017). *Social skills improvement system social emotional learning edition rating forms*. Pearson Assessments.
- Gresham, F. M., Elliott, S. N., Byrd, S., Wilson, E., & Cassidy, K. (2018). Cross-informant agreement of children's social emotional skills: An investigation of ratings by teachers, parents, and students from a nationally representative sample. *Psychology in the Schools*, 55, 208–223. <https://doi.org/10.1002/pits.22101>
- Gresham, F. M., Elliott, S. N., Metallo, S., Byrd, S., Erickson, M., & Altman, R. A. (2020). Psychometric fundamentals of the social skills improvement system social learning edition rating forms. *Assessment for Effective Intervention*, 45(3), 194–209. <https://doi.org/10.1177/1534508418808598>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- McKown, C. (2019). *Assessing students' social emotional learning: A guide to meaningful measurement*. Norton & Company.
- Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *The Journal of Applied Psychology*, 95(4), 728–743. <https://doi.org/10.1037/a0018966>
- Moulton, S., von der Embse, N., Kilgus, S., & Drymond, M. (2019). Building a better behavior progress monitoring tool using maximally efficient items. *School Psychology (Washington, DC)*, 34(6), 695–705. <https://doi.org/10.1037/spq0000334>

- Muthén, L. K., & Muthén, B. (1998–2019). Mplus. The comprehensive modelling program for applied researchers: User's guide, 5.
- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011/2014). *The Devereux student strengths assessment – mini (DESSA-Mini): Assessment, technical manual, and user's guide*. Apperson.
- National Institute of Mental Health. (2020). *Child and adolescent mental health*. <https://www.nimh.nih.gov/health/education-awareness/shareable-resources-on-child-and-adolescent-mental-health=shareNIMH>.
- 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. <https://doi.org/10.1016/j.jval.2014.10.005>
- Petersen, K. J., Humphrey, N., & Qualter, P. (2020). Latent class analysis of mental health in middle childhood: Evidence for the dual-factor model. *School Mental Health*, 12(4), 786–800. <https://doi.org/10.1007/s12310-020-09384-9>
- Reeve, B. B., Hays, R. D., Chang, C. H., & Perfetto, E. M. (2007). Applying item response theory to enhance health outcomes assessment. *Quality of Life Research*, 16(S1), 1–3. <https://doi.org/10.1007/s11136-007-9220-6>
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). American Guidance Service.
- Reynolds, C., & Kamphaus, R. W. (2015). *Behavior assessment system for children* (3rd ed.). Minneapolis, MN: NCS Pearson, Inc.
- Ross, K. M., & Tolan, P. (2017). Social and emotional learning in adolescence: Testing the CASEL model in a normative sample. *Journal of Early Adolescence*, 40(1), 1–30. <https://doi.org/10.1177/0272431617725198>.
- Rupp, S., Elliott, S. N., & Gresham, F. M. (2018). Assessing elementary students' bullying and related social behaviors: Cross-informant consistency across school and home environments. *Children and Youth Services Review*, 93, 458–466. <https://doi.org/10.1016/j.childyouth.2018.08.028>
- Saeki, E., Jimerson, S. R., Earhart, J., Hart, S. R., Renshaw, T., Singh, R. E., & Stewart, K. (2011). Response to intervention (RtI) in the social, emotional, and behavioral domains: Current challenges and emerging possibilities. *Contemporary School Psychology*, 15, 43–52.
- Salvia, J., Ysseldyke, J. E., & Witmer, S. (2016). *Assessment in special and inclusive education* (13th ed.). Cengage Learning.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Psychometrika*, 34(S1), 1–97. <https://doi.org/10.1007/BF03372160>
- U.S. Department of Education Office for Civil Rights. (2016). *Civil rights data collection: A first look*. <https://www2.ed.gov/>

AUTHOR BIOGRAPHICAL STATEMENTS

Stephen Nelson Elliott, PhD, is an educational psychologist and the Mickelson Foundation Professor in the Sanford School of Social and Family Dynamics at Arizona State University. Steve has collaborated with over 100 colleagues and students to author research articles, books, and chapters focusing on ways to improve the social and academic performance of students at risk or with disabilities. He has been recognized as a Fellow in the American Educational Research Association and Senior Scientist in the American Psychological Association. He served as Editor of School Psychology Review from 1984–1990.

Pui-Wa Lei, PhD, is a professor in the Department of Educational and School Psychology and Special Education at the Pennsylvania State University. She specializes in Educational Measurement and Statistics. Her research interests are in methodological issues of multivariate statistical analyses, particularly structural equation modeling and multilevel modeling, and in applications of these modeling techniques and item response theory. She currently serves on the editorial board of the Journal of Educational Psychology.

Christopher J. Anthony, PhD, is an assistant professor in the School of Special Education, School Psychology, and Early Childhood Studies in the College of Education at the University of Florida. His research focuses broadly on improving the assessment of positive student competencies, especially academic enablers and social and emotional learning. He currently serves on the editorial boards of the Journal of School Psychology and School Psychology Review.

James Clyde DiPerna, PhD, is Professor and Director of the School Psychology Program at The Pennsylvania State University. His research program focuses on the development of assessments and interventions to enhance students' social, emotional, and academic competence in school settings.