Hammill Institute on Disabilities

Topics in Early Childhood Special Education
2015, Vol. 34(4) 234–249

DOI: 10.1177/0271121414544801
tecse.sagepub.com
SAGE

The Student Risk Screening Scale for Early Childhood: An Initial Validation Study

Kathleen Lynne Lane, PhD[1], Wendy Peia Oakes, PhD[2], Holly Mariah Menzies, PhD[3], Rebecca Major, MEd[4], Laurie Allegra, MEd[5], Lisa Powers, PhD[6], and Chris Schatschneider, PhD[7]

Abstract

We report findings of two exploratory validation studies of a revised instrument: the *Student Risk Screening Scale for Early Childhood version* (SRSS-EC). The SRSS-EC was modified to reflect characteristics of externalizing and internalizing behaviors manifested by preschool-age children. In Study 1, we explored the reliability of SRSS-IE (*Student Risk Screening Scale–Internalizing and Externalizing*) 14 scores by examining item-level data, internal consistency, and factor structure using exploratory factor analysis with 489 preschool-age students. Results supported retention of 11 items, resulting in the SRSS-EC 11 (with SRSS-EC E7 externalizing subscale and SRSS-EC I4 internalizing subscale). Next, we established convergent validity of the SRSS-EC 11 with two well-established screening tools: the Strengths and Difficulties Questionnaire and the Early Screening Project. In Study 2, we reported findings of a confirmatory factor analysis and a subsequent convergent validity study with 737 preschool-age students to determine generalizability. We offered limitations and future directions.

Keywords

systematic screening, internalizing and externalizing behavior, preschool, early childhood

Systematic screening tools for behavior are being introduced within multi-tiered systems of support to accurately detect and assist school leadership teams in detecting school-age students—preschool, elementary, middle, and high school—with behavioral challenges at the earliest sign of concern. This is encouraging given that behavioral challenges include a breadth of concerns ranging from internalizing (e.g., anxiety, social withdrawal, and loneliness) to externalizing behaviors (e.g., noncompliance and aggression) and include a greater percentage of children than one might expect. Prevalence estimates suggest that 20% of school-age youth exhibit mild-to-severe forms of such emotional and behavioral disorders (EBDs), and 12% exhibit moderate-to-severe forms (Forness, Freeman, Paparella, Kauffman, & Walker, 2012).

Multi-tiered systems such as The Pyramid Model (Hemmeter & Conroy, 2012), Positive Behavioral Interventions and Supports (PBIS; Dunlap, Sailor, Horner, & Sugai, 2009), and Comprehensive, Integrated Three-Tiered (CI3T) models of prevention (Lane, Kalberg, & Menzies, 2009) provide a prevention framework in which data from systematic screening tools can be used to accurately detect and respond to students' needs. These models share a similar structure: primary prevention (Tier 1) for every student including schoolwide programming for students across all domains (e.g., academics, behavioral, social skills), with additional supports and instruction at Tier 2 (targeted intervention) and Tier 3 (intensive intervention) for students requiring assistance beyond Tier 1 efforts (Fox, Carta, Strain, Dunlap, & Hemmeter, 2010).

Proactive, systematic screening practices are a critical component of these models to ensure that each student—from preschool through high school—is considered for additional needs at multiple times (fall, winter, and spring)

[1]University of Kansas, Lawrence, USA
[2]Arizona State University, Mesa, USA
[3]California State University, Los Angeles, USA
[4]Warren, Kentucky
[5]Nashville, TN
[6]Special School District of St. Louis County, MO, USA
[7]Florida State University, Tallahassee, USA

Corresponding Author:
Kathleen Lynne Lane, Professor, Department of Special Education, University of Kansas, 1122 West Campus Road, JRP Room 720, Lawrence, KS 60045, USA.
Email: Kathleen.Lane@ku.edu

over the course of each academic year (Lane & Walker, in press). Ideally, such screenings enable early detection, with early referring to two facets: (a) early in the development of learning, behavior, or social concerns that are developing and (b) early in a child's educational career, when they are most responsive to intervention efforts.

While we recommend social and emotional screening practice for very young children, the preschool years offer the first formalized school experience for students and an ideal early opportunity for addressing the social and emotional needs of young children with and at risk for EBD when these behavioral patterns are most malleable (Lane & Walker, in press). Prevalence estimates of EBDs in preschool-age children illuminate the need for early detection efforts at this level. In a review of the prevalence literature on emotional and behavioral psychiatric disorders in preschool-age children, Egger and Angold (2006) reported on the five most common types of disorders—attention deficit hyperactivity, oppositional defiant (ODD), conduct disorders, anxiety, and depressive, indicating a prevalence rate of 19.5% for preschool-age students (age 2–5). These prevalence and comorbidity rates were similar to those reported by Forness et al. (2012) for older children and youth as noted previously. Comorbidity rates continue to be particularly high between depressive disorders and ODD (Wichstrom et al., 2012), indicating the importance of screening for internalizing *and* externalizing behaviors. As such, systematic screening should begin early—in the preschool years. National recognition and support for early detection of social and emotional delays exists and is a priority for preschool-age children, with 70% of states recommending screening practices for young children (Cooper & Vick, 2009).

To meet this need, a growing number of tools exist for this proactive purpose: early detection to respond to individual student needs. Behavior screening tools are available across the PK–12 grade span, with some including multiple-informant versions. Examples include the BASC-2 Behavioral and Emotional Screener (BASC-2 BESS; Kamphaus & Reynolds, 2007), Early Screening Project (ESP; H. M. Walker, Severson, & Feil, 1995), Social Skills Improvement System Performance Screening Guide (SSiS-PSG; Elliott & Gresham, 2007), Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), *Student Risk Screening Scale* (SRSS; Drummond, 1994), and Systematic Screening for Behavior Disorders (SSBD; H. M. Walker & Severson, 1992). Descriptions of each are beyond the scope of this article but can be found elsewhere (Feeney-Kettler, Kratochwill, Kaiser, Hemmeter, & Kettler, 2010; Kettler, Glover, Albers, & Feeney-Kettler, 2014; Lane, Menzies, Oakes, & Kalberg, 2012), yet we note that the BASC-2 BESS, ESP, SSiS-PSG, and SDQ include options for use at the preschool level.

The Ages and Stages Questionnaire: Social Emotional (ASQ:SE; Squires, Bricker, & Twombly, 2002) is another commonly used screening tool, which allows for developmental screening of young children (6 to 66 months) in the social and emotional domain by parents, caregivers, and teachers. This well-established tool is a reliable, valid, low-cost measure and often used in preschool centers. Yet, in a recent review of screening tools for young children, the ASQ:SE was not recommended as a classwide screening tool due to concerns with feasibility, requiring 10 to 15 min of teacher time per student to complete this tool (Feeney-Kettler et al., 2010).

When selecting a screening tool, school- and district-level leadership teams must consider multiple questions regarding psychometric rigor, feasibility, and utility. For example, it is important to ensure that (a) the selected tool is reliable and valid for the student population, (b) the tool has utility for early detection of potential problems to guide the intervention decisions as well as monitoring program effectiveness, (c) the tool reports information from observations of children in authentic settings, and (d) requisite resources are available and/or secured to ensure proper implementation (e.g., time and budget for preparation, administration, scoring; Feeney-Kettler et al., 2010; Neisworth & Bagnato, 2005). Fortunately, there are options for teams with limited monetary resources who are interested in implementing screening practices. The SRSS and SDQ are free-access, psychometrically sound, widely used tools and the ESP is a low-cost tool with specific subscales to assess internalizing and externalizing concerns (descriptions to follow).

An Expanded Scope for the SRSS

Of these three tools, the SRSS requires the least amount of teacher time (10 min to screen a whole class)—another precious commodity. Yet, the SRSS was not initially developed to detect internalizing as well as externalizing behaviors and it was not designed for use at the preschool level. To address this first limitation, the SRSS was modified recently to expand the tool's ability to detect students with internalizing as well as externalizing behaviors. In the initial validation study, Lane, Oakes, et al. (2012) examined an adapted tool—the *Student Risk Screening Scale–Internalizing and Externalizing* (SRSS-IE)—which was modified to include seven additional items to detect internalizing behaviors. In this study of 2,460 K–6th grade students, results offered initial evidence to suggest that five of the new items should be retained to detect internalizing behaviors. Results indicated an internal consistency estimate of .83 for the full scale (the original seven items and an additional five items to measure internalizing behaviors) and a two-factor structure. Findings also offered evidence of convergent validity between SRSS-IE scores and two existing measures—the SDQ and SSBD. Convergent validity was established between SRSS-IE and SDQ scores, with

all scales statistically significantly and moderately-to-highly correlated. In comparing SSBD and SRSS-IE scores, a receiver operating characteristic (ROC) curve analysis provided initial evidence for using SRSS-IE scores to detect externalizing and/or internalizing behaviors.

Next, Lane, Menzies, Oakes, Lambert, and colleagues (2012) conducted two additional studies in a southeastern state with elementary students ($N = 982$, $N = 1{,}079$) to offer further evidence for the psychometric adequacy of the SRSS-IE. Internal consistency estimates and factor structure results were consistent with the Lane, Oakes, et al. (2012) findings. Again convergent validity was established between SSBD and SRSS-IE scores, improving chance estimates 34% to 42%.

Following confirmation of the utility of SRSS-IE at the elementary level, Lane, Oakes, Carter, Lambert, and Jenkins (2013) conducted an initial look at the reliability and validity of the SRSS-IE for use at the middle school level. In this sample of 937 students in a southeastern state, alpha coefficients were .82 and a two-factor structure established, providing preliminary evidence for the use of the SRSS-IE with middle-school-age students.

Collectively, these first three studies of the SRSS-IE provided evidence for the reliability and validity of this tool for detecting students with and at risk for the two major disorders of childhood: externalizing and internalizing behaviors. To address the absence of a free-access and very brief screening tool for use at the preschool level, Lane and Menzies (in press) developed a downward extension of the SRSS-IE for preschool students—the *Student Risk Screening Scale–Early Childhood version* (SRSS-EC) examined for the first time in this article. In the two studies that follow, we report initial results for this preschool screening measure—an adapted version of the SRSS-IE intended to reflect the developmental considerations of the preschool years.

Purpose

In this article, we report findings of two validation studies of the adapted instrument: the SRSS-EC as applied with preschool-age students attending one of three early childhood centers in a Midwestern state. This adapted tool—the SRSS-EC—is described in detail in the method. The SRSS-EC was modeled after the SRSS in format and SRSS-IE in content. The intent was to create a downward extension of these tools for use with preschool students, offering elementary schools with PK classrooms comparable schoolwide approaches to detect PK students who may need additional supports. The initial 14 items were gleaned from the current literature base (e.g., screening studies conducted with preschool students; for example, Coplan & Rubin, 1998; Egger & Angold, 2006), a review of screening tools available for the preschool level, and professional experiences with students with EBD. The teacher-completed SRSS-EC 14 requires the lead preschool teachers to rate each item on a 4-point Likert-type scale ranging from 0 (*never*) to 3 (*frequently*), the same scale developed for the SRSS.

In the first study, we used classical test theory to explore the initial evidence for the reliability and validity of the SRSS-EC for use with preschool-age students. We examined the reliability of the SRSS-EC by analyzing item-level data, internal consistency estimates, and factor structure using an exploratory factor analysis (EFA) with data collected from two early childhood centers ($N = 489$). Then, we explored convergent validity of the SRSS-EC with two established tools: the SDQ and ESP. In the second study, we conducted a confirmatory factor analysis (CFA) and a subsequent validity study with 737 children from a third early childhood center in the same locale to determine the generalizability of the findings reported in Study 1.

Study 1: Method

Participants and Setting

Participants were 489 children (279 [57.06%] boys) attending one of two culturally and economically diverse public early childhood centers in a Midwestern state, both classified as Suburban: Large according to locale codes (National Center for Education Statistics, n.d.). Neither center received Title 1 services. Each child was rated by their classroom teacher ($N = 22$) on three screening tools: the SRSS-EC, SDQ, and Stages 1 and 2 of the ESP. Most children were White (88.73%; $n = 433$), with 31.08% ($n = 152$) receiving special education services (see Table 1). Students ranged in age from 3.03 to 5.57 years ($M = 4.33$, $SD = 0.64$).

Nine teachers were from Center A and 13 from Center B. In Center A, teachers rated between 17 and 35 students (as some teachers taught morning and afternoon classes; $M = 22.44$, $SD = 5.66$). All teachers at Center A were White and female, with the exception of one Asian female. Teachers ranged in age from 24 to 55 years ($M = 38.00$, $SD = 10.79$), with 1 to 23 years of teaching experience as an early childhood teacher ($M = 10.89$, $SD = 7.51$). In Center B, teachers rated between 4 and 39 students ($M = 22.44$, $SD = 5.66$). All teachers at Center B were White females ranging in age from 24 to 62 years ($M = 37.23$, $SD = 12.46$). Teachers had 1 to 16 years of teaching experience as an early childhood teacher ($M = 7.15$, $SD = 4.69$).

Procedures

This study was designed in partnership with district leaders in a Midwestern state who were interested in exploring systematic screening tools for use at the preschool level. Namely, district leaders were looking for a psychometrically

Table 1. Student Characteristics.

	Study 1			Study 2
	Center A, *N* = 202	Center B, *N* = 287	Total *N* = 489	Center C, *N* = 737
Gender (%, *n*)				
Boys	53.96 (109)	59.23 (170)	57.06 (279)	57.82 (425)
Girls	46.04 (93)	40.77 (117)	42.94 (210)	42.18 (310)
Ethnicity (%, *n*)				
White	89.55 (108)	88.15 (253)	88.73 (433)	88.57 (651)
Black	4.48 (9)	3.48 (10)	3.89 (19)	1.90 (14)
Hispanic	0.00 (0)	0.35 (1)	0.20 (1)	2.86 (21)
Asian	3.48 (7)	5.23 (15)	4.51 (22)	6.12 (45)
Native American	0.00 (0)	0.00 (0)	0.00 (0)	0.14 (1)
Other	1.00 (2)	1.05 (3)	1.02 (5)	0.41 (3)
Biracial	1.49 (3)	1.74 (5)	1.64 (8)	0.00 (0)
Special education services at the time of screening (%, *n*)	24.75 (50)	35.54 (102)	31.08 (152)	23.78 (175)
Intellectual disability	0	0	0	2
Speech impairment	10	8	18	0
Language impairment	0	2	2	0
Emotional disturbance	0	0	0	1
Autism	0	6	6	0
Orthopedic impairment	1	0	1	0
Deafness	0	0	0	1
Blind	0	0	0	2
Functional delay	0	0	0	157
Developmental delay	39	85	124	1
Traumatic brain injury	0	0	0	10
Speech and language impairment	0	1	1	0
Age (*M, SD*)	4.29 (0.61)	4.35 (0.65)	4.33 (0.64)	4.50 (0.61)
Class time (%, *n*)				
a.m.	18.32 (37)	48.43 (139)	35.99 (176)	47.49 (350)
p.m.	24.26 (49)	50.17 (144)	39.47 (193)	25.51 (188)
All day	0.00 (0)	0.00 (0)	0.00 (0)	26.87 (198)
Information not provided	57.43 (116)	1.39 (4)	24.54 (120)	0.14 (1)

Note. Percentages are based on the number of participants who completed the item. Children ranged in age from 3.03 to 5.58 years.

strong and socially valid screening tool that would be feasible for use in terms of cost and time required for preparation, administration, scoring, and interpretation. The centers were using the ESP as part of regular practices; yet, they were struggling with issues of scoring and interpretation as decision-making process for the ESP does not include the same level of precision (e.g., decision trees) offered by the SSBD (which had been used at feeder elementary schools). After obtaining university and district approvals, three early childhood centers were invited to participate in this unfunded validation study. All centers expressed interest in participating. Study 1 involved Centers A and B.

The research team prepared and mailed a packet to each early childhood center that included separate packets for each lead teacher who worked with students ages 3 to 6. Packets included the following: (a) two copies of the consent form—one to sign and return and the other to retain for their records, (b) a teacher-completed demographic form (e.g., gender, years of experience, highest degree obtained), (c) copies of the three screening tools (SRSS-EC, SDQ, and ESP), and (d) a student demographic sheet (e.g., age, gender, birth month and year, disability status). Student names were not used on any of the study measures, instead teachers were provided with a master sheet by the school administrator where they assigned a number (1–20) to each student. Teachers maintained those master sheets which were not shared with the research team. Teachers used the numbers assigned as the identifier for each student on screening and demographic information.

The primary investigators (PIs) conducted consenting meetings remotely, with district partners on-site to facilitate technology. At Centers A and B, all teachers met in one room. The PI provided an overview of the rationale for the study using the consent form, noting potential risks (e.g.,

loss of time) and benefits (e.g., contributing to the development of a user-friendly, free-access screening tool), and addressed all questions. All teachers at Center A ($n = 9$) and Center B ($n = 13$) elected to participate and completed the demographic forms and screening tools at that time, requiring about 2 hr. The on-site district partner and her staff collected completed packets from each teacher and mailed them with a tracking number to the PIs where the research team checked in the data. Research assistants entered data into spreadsheets, checking 25% for reliability of data entry (>95% accuracy). Consented teachers who returned a completed packet (all three screeners and the demographic sheets) were entered into a drawing to receive a US$100 Visa gift card, with one gift card awarded at each center.

Measures

Measures included the SRSS-EC, SDQ, and ESP, each completed in succession and described below. Measures were not counterbalanced for completion, a noted limitation.

SRSS-EC. The SRSS-EC was modeled after the SRSS, a free-access, brief screening tool developed to detect elementary-age students with antisocial behaviors. The SRSS includes seven items—steal; lie, cheat, sneak; behavior problem; peer rejection; low academic achievement; negative attitude; and aggressive behavior—which are rated on a 4-point Likert-type scale (0 = *never*, 1 = *occasionally*, 2 = *sometimes*, 3 = *frequently*). Scores are totaled and evaluated based on three categories: low (0–3), moderate (4–8), and high risk (9–21). Cut scores were developed by Drummond (1994) for use at the elementary level. The SSRS was then adapted to expand the scope of this tool to detect students with internalizing as well as externalizing behaviors by adding items to reflect characteristic behavior patterns of internalizing behavior (five retained items: emotionally flat; shy, withdrawn; sad, depressed; anxious; lonely). The goal of developing the SRSS-EC was to create a comparable but developmentally appropriate measure in terms of design as well as ease of preparation, completion, and scoring.

To develop the SRSS-EC, we invoked the same process used to develop the SRSS-IE. We examined systematic literature reviews (e.g., Costello, Erkanli, & Angold, 2006); descriptive, longitudinal studies conducted of behavioral characteristics of preschool-age students; and validation studies for existing screening tools available for use at the preschool level such as the ESP, SDQ, SSiS-PSG, and BASC-2 BESS. Also, we reviewed the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association, 2000) and technical manuals for existing rating scale tools (e.g., SSiS Rating Scale; Gresham & Elliott, 2008) and BASC-2 behavior rating scales. We reviewed these key sources containing the most rigorous information available in defining developmentally appropriate, core characteristic behavior patterns of internalizing and externalizing constructs. We included items believed to reflect common behavior representing both major disorders of childhood.

SDQ. The SDQ is a free-access, universal screening tool used to assess students ages 3 to 17 years. Parent, teacher, and student self-report (age 11–17) versions are available to assess behavioral skills across five-factor analytically derived scales: emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behaviors. The teacher version used in this study included 25 items rated on a 3-point, Likert-type scale (*not true* = 0, *somewhat true* = 1, *certainly true* = 2), with negatively worded items reverse scored. A total difficulties score is computed by adding scores from each domain except the prosocial scale. Cut scores yield three categories: Normal, Abnormal, and Borderline. The SDQ teacher version has adequate test–retest stability correlations (.65–.82) and internal consistency (α = .70–.87; Goodman, 2001).

ESP. The ESP is an empirically validated, multiple-gating screening tool used to identify children aged 3 to 5 who may be at heightened risk for externalizing and internalizing behaviors. In Stage 1, teachers consider all students in their class and nominate students whose characteristic behaviors are most consistent with the descriptors for externalizing ($n = 5$) and internalizing ($n = 5$) behaviors. Next, teachers rank order nominated students from most to least like each category, creating two mutually exclusive lists. The top three students on each list pass through Gate 1 to Stage 2. In Stage 2, teachers completed five additional scales: *Critical Events Index* (CEI), *Adaptive Behavior Scale* (CFI-A), *Maladaptive Behavior Scale* (CFI-M), *Aggressive Behavior Scale* (ABS; only for students in the externalizing category), and *Social Interaction Scale* (SIS; only for students nominated in the internalizing category). These checklists further describe the type and intensity of internalizing and externalizing behaviors that may place students at increased risk for developing EBD. The CEI examines high intensity, low frequency behaviors (16 items) such as steals and sets fires. The remaining four scales represent lower intensity and higher frequency behaviors. Scores on the index and scales are compared with a normative sample of students by age and gender when determining risk. Scores place students into the following categories: low risk, at risk, high risk, and extreme risk for boys and girls for CEI, CFI-A, CFI-M, and ABS subscale and ranging between low risk and at risk for boy and girls for the SIS subscale. The measure takes 30 to 40 min to evaluate a classroom of 20 preschool students. Test–retest reliability ranges from .74 (CEI) to .90 (ABS) for Stage 2 (Feil, Severson, & Walker, 1995).

Statistical Analysis

The analytic plan employed was grounded in classical test theory (CTT; Cronbach & Shavelson, 2004), EFAs, and graphics modeled after L. S. Walker, Beck, Garber, and Lambert (2009). Validity was assessed by examining convergent validity between SRSS-EC (total and subscale scores) and SDQ and ESP scores. First, we examined reliability of SRSS-EC scores. We evaluated item-level descriptive statistics to identify items with a possible "floored" effect. Next, we conducted an EFA of the 13-item instrument (SRSS-EC 13) using Velicer's *Minimum Average Partial test* (MAP; Velicer, 1976) and a parallel analysis test (Horn, 1965) to determine the number of factors to retain. Then, we computed alpha coefficients, with $\alpha = .80$ considered adequate (Nunnally & Bernstein, 1994).

A range of reliability indicators were reviewed to detect potentially problematic items which, if eliminated, might improve scale reliability. Consistent with Lane, Menzies, Oakes, Lambert, et al. (2012) and Lane, Oakes, et al. (2012) data analytics, empirical and theoretical considerations regarding externalizing and internalizing constructs were used to determine which items to retain. Figures of merit were treated as guidelines rather than pass–fail criteria. Results were analyzed in tandem with a theoretical perspective to retain psychometrically sound items that were also consistent with characteristic behaviors reflecting the construct of interest (content validity).

Second, we examined convergent validity of the SRSS-EC using two criterion measures: the SDQ and ESP. We examined concurrent validity of the SRSS-EC through correlations with existing, psychometrically sound instruments. We examined convergent validity between SRSS-EC E7, SRSS-EC I4, and SRSS-EC 11 and (a) SDQ scores (total and subscale scores) using Pearson's correlation coefficients as the full sample included continuous scores on both measures and (b) ESP subscale categories using Spearman's correlation coefficients to accommodate the categorical ESP variables. Consistent with validity studies of the SRSS and SRSS-IE at the elementary and middle school level, we hypothesized that the SRSS-EC E7 scores would have higher correlations with the conduct problems and hyperactivity scores and low correlations with emotional symptoms and peer problems and the converse with SRSS-EC I4 scores. We also hypothesized that the SRSS-EC 11 total score would have a significant correlation with the CEI scores, which includes both externalizing and internalizing behaviors. We also anticipated that (a) SRSS-EC E7 scores would have a large correlation with CFI-M and ABS scores as we believed both to be reflective of antisocial tendencies, and this measure was only completed for children nominated in the externalizing category and (b) SRSS-EC I4 scores would have large correlations with SIS scores as struggles with peers are characteristic of students with internalizing issues, and this measure was only completed for children nominated in the internalizing category. Correlation coefficients > .50 are considered large (Hopkins, 2002).

Study 1: Results and Summary

In the results, we refer to the initially developed 14 items as the SRSS-EC 14. Items 1 to 8 were proposed to measure externalizing behaviors, and Items 9 to 14 were proposed to measure internalizing behaviors. The SRSS-EC14 had no missing data for the 489 participants.

Reliability

Item-level analyses. First, we examined the psychometric properties of the SRSS-EC14 (see Table 2). CTT indicates that sound items should be free of floor and ceiling effects, which restrict variance and yield excessive skewness and kurtosis. CTT also indicates that items should have high item-total correlations (>.35) to afford high internal consistency (high α reliability; L. S. Walker et al., 2009). Teachers rated each of the 14 items on a Likert-type scale from 0 to 3 resulting in mean scores ranging from 0.13 for *steals* to 1.15 for *plays alone*, with several items having low mean scores. For example, the item *steals* had 91.21% ($n = 446$) zeros, suggesting that this item had a very low base rate of occurrence and contributed limited information about most children as it was rarely endorsed (L. S. Walker et al., 2009). In Table 2, mean values below 0.20 appear in bold to identify items with possible flooring, with only *steals* falling below the 0.20 criterion. The *steals* item also had another indicator of concern, with a kurtosis score of 16.11 (>15 criteria). Low mean values of <1.00, with the exception of *plays alone*, were expected, given that the sample was all students attending these two centers, not only those identified as having concerns. We elected to remove the *steals* item as item-level analyses suggested the noted concerns and stealing (e.g., taking an object such as a toy or food without permission) may not be a developmentally appropriate item to include at the preschool level.

Factor structure. Responses to the initial 13 items (after eliminating *steals* a priori) were entered into an EFA using squared multiple correlations as prior communality estimates (see Table 3). The principal factor method was used to extract factors, followed by a promax (oblique) rotation as we expected externalizing and internalizing constructs to be moderately correlated (Achenbach, 1991). Three factors were retained for rotation according to evidence from both *Velicer's MAP* test and parallel analyses test, which were computed using a program written by O'Connor (2000).

When interpreting the rotated factor pattern, an item was considered to load on a given factor when the factor loading

Table 2. Psychometric Properties of the 14 SRSS-EC Items.

	Study 1					Study 2				
	Psychometrics					Psychometrics				
Time and variable	*M* < 0.20	*SD*	Skewness > 4	Kurtosis > 15	Eliminate	*M* < 0.20	*SD*	Skewness > 4	Kurtosis > 15	Eliminate
1. Tantrums	0.59	0.98	1.45	0.67		0.39	0.82	2.08	3.20	
2. Active, restless	0.93	1.12	0.76	−0.94		0.80	1.11	1.04	−0.44	
3. Steals	**0.13**	0.47	3.95	**16.11**	*	**0.07**	0.32	**4.87**	**26.38**	*
4. Rejected by peers	0.44	0.78	1.72	2.07		0.27	0.61	2.47	5.96	
5. Ignores teacher and class rules	0.87	1.04	0.82	−0.66		0.67	0.98	1.23	0.23	
6. Negative attitude	0.52	0.88	1.52	1.10		0.43	0.76	1.79	2.42	
7. aggressive behavior	0.44	0.84	1.84	2.30		0.34	0.74	2.16	3.70	
8. Lies	0.20	0.55	3.03	9.24		**0.14**	0.45	3.35	11.48	
9. Shy, timid	0.83	1.07	0.99	−0.40		0.81	0.98	0.93	−0.29	
10. Sad, tearful	0.46	0.82	1.70	1.88		0.46	0.72	1.47	1.47	
11. Worried, fearful	0.42	0.81	1.87	2.39		0.35	0.66	1.89	3.03	
12. Talks with other children (*R*)	0.67	0.95	1.18	0.17		0.68	0.97	1.26	0.39	
13. Plays alone	1.15	1.05	0.48	−0.97		1.09	0.97	0.63	−0.54	
14. Physical complaints (e.g., stomach hurts)	0.21	0.59	3.09	9.55		0.23	0.57	2.70	7.41	

Note. Bolded numbers indicated those values that may be problematic as they exceed the established threshold noted at the top of each column. Items with an asterisk were removed from the final version. In Studies 1 and 2, all items ranged from 0 to 3; Study 1, *N* = 489; Study 2, *N* = 738, except for Item 2, which had one missing observation. SRSS-EC = Student Risk Screening Scale for Early Childhood version.

Table 3. Exploratory Factor Analysis SRSS-EC 11: Rotated Factor Pattern Matrix, Factor Structure Matrix, and Communality Estimates: Study 1.

	SRSS-EC 11				
	Rotated factor pattern (Promax)		Factor structure (correlations)		
Item	1	2	1	2	h^2
1. Tantrums	.68*	.17	.73*	.37	.56
2. Active, restless	.79*	−.10	.76*	.13	.58
4. Rejected by peers	.56*	.10	.59*	.27	.35
5. Ignores teacher and class rules	.83*	.10	.82*	.20	.67
6. Negative attitude	.57*	−.04	.64*	.42*	.47
7. Aggressive behavior	.80*	−.12	.77*	.12	.60
8. Lies	.40*	−.10	.37	.01	.14
9. Shy, timid	−.25	.57*	−.08	.50*	.30
10. Sad, tearful	.13	.73*	.35	.77*	.61
11. Worried, fearful	−.05	.76*	.17	.74*	.55
14. Physical complaints (e.g., stomach hurts)	.17	.49*	.31	.54*	.32

Note. Items with an asterisk were those with factor loadings ≥.40. Promax rotation permits Factor 1 and Factor 2 to correlate with each other. h^2 are communality estimates, the sum of the squared factor loadings on all factors for a given variable to show shared variance. Low communality indicates an item's uniqueness—that the factors do not explain it well. SRSS-EC = Student Risk Screening Scale for Early Childhood version.

was ≥ .40 on the given factor and <.40 on the remaining factor. Applying these criteria, *tantrums* (1), *active, restless* (2), *rejected by peers* (4), *ignores teacher and class rules* (5), *negative attitude* (6), *aggressive behavior* (7), and *lies* (8) loaded on the externalizing factor (Factor 1); *shy timid* (9), *sad, tearful* (10), *worried, fearful* (11), and *physical complains* (14) loaded on the internalizing factor (Factor 2); and items *talks with other children* (12R) and *plays alone* (13) loaded on a third factor (Factor 3). We examined the rotated factor patterns which indicated the unique

Table 4. Cronbach's Coefficient Alphas With Standardized Variables: Studies 1 and 2.

	Study 1		Study 2	
	r with total (<.35)	α^a	*r* with total (<.35)	α^a
Item	Externalizing items (7)		Externalizing items (7)	
		.85		.88
1. Tantrums	.64	.82	.67	.86
2. Active, restless	.68	.81	.72	.85
4. Rejected by peers	.55	.83	.57	.87
5. Ignores teacher and class rules	.77	.80	.82	.84
6. Negative attitude	.60	.83	.68	.86
7. Aggressive behavior	.69	.81	.75	.85
8. Lies	**.32**	.87	.43	.89
	Internalizing items (4)		Internalizing items (4)	
		.73		.60
9. Shy, timid	.37	.75	.31	.58
10. Sad, tearful	.63	.60	.51	.42
11. Worried, fearful	.66	.60	.53	.41
14. Physical complaints	.43	.72	**.20**	.66
	11 items		11 items	
		.81		.83
1. Tantrums	.64	.78	.62	.80
2. Active, restless	.56	.79	.64	.80
4. Rejected by peers	.52	.79	.57	.80
5. Ignores teacher and class rules	.65	.78	.74	.79
6. Negative attitude	.64	.78	.69	.79
7. Aggressive behavior	.56	.79	.65	.80
8. Lies	**.25**	.82	.38	.82
9. Shy, timid	**.11**	.83	**.01**	.85
10. Sad, tearful	.52	.79	.56	.81
11. Worried, fearful	.39	.81	.33	.83
14. Physical complaints	.42	.80	**.32**	.83

Note. Subsequent values refer to the alpha values if the item had been deleted from the scale. Study 1, *N* = 489; Study 2, *N* = 737.
[a]In the columns labeled, α, the first alpha value is the overall alpha level.

contribution each factor makes to the unique variance of the variable as well as the factor structure matrix which indicated the correlation between a given factor and a variable (Hatcher, 1994). The pattern matrix (standardized regression coefficients) was analyzed to ascertain which groups of variables measure a specific factor (externalizing or internalizing) with the intent of interpreting the meaning of the factor. Next, the factor structure matrix was reviewed to determine the bivariate relations between factors and variables.

Based on the rotated factor pattern and factor structure matrices, all seven items (1, 2, 4, 5, 6, 7, and 8) were retained on the externalizing factor, and the four items (9, 10, 11, and 14) were retained on the internalizing factor. We eliminated items constituting the proposed third factor (12R and 13) as a minimum of three items are required to constitute a factor. The variance explained by Factor 1 (Externalizing) 30.00% and variance explained by Factor 2 (Internalizing) increased to 16.91% (interfactor correlation = .29), for a total variance of 46.99%.

Internal consistency. Next, we computed Cronbach's alpha coefficients for the seven items in Factor 1 proposed to measure characteristic overt and covert externalizing behaviors (SRSS-EC E7). The alpha coefficient was .85, above the established popular criterion of α = .80 (Nunnally & Bernstein, 1994). Item-total correlations ranged from .32 for *lies* (a more covert behavior, just slightly below the .35 criteria for *r* with total score) to .77 for *ignores teacher and class rules* (see Table 4). In Table 4, we bolded item-total

Table 5. Convergent Validity: SRSS-EC E7, SRSS-EC I4, and SRSS-EC 13 With the SDQ.

	Study 1 (*N* = 489)					Study 2 (*N* = 737)				
			SRSS-EC E7	SRSS-EC I4,	SRSS-EC 11			SRSS-EC E7	SRSS-EC I4,	SRSS-EC 11
SDQ scales	*M*	*SD*	*r* (*n*)	*r* (*n*)	*r* (*n*)	*M*	*SD*	*r* (*n*)	*r* (*n*)	*r* (*n*)
Total	8.97	7.42	.79 (481)	.42 (481)	.82 (481)	7.24	7.54	.83 (726)	.40 (726)	.84 (726)
Emotional symptoms	1.54	2.15	.21 (488)	.76 (488)	.50 (488)	1.25	1.79	.35 (732)	.66 (732)	.55 (732)
Conduct problems	1.60	2.32	.81 (487)	.08[a] (487)	.69 (487)	1.31	2.28	.85 (729)	.15 (729)	.77 (729)
Hyperactivity	3.57	3.42	.73 (487)	.13[b] (487)	.65 (487)	2.96	3.34	.77 (733)	.21 (733)	.72 (733)
Peer problems	2.32	2.35	.44 (486)	.35 (486)	.50 (486)	1.72	2.08	.53 (733)	.35 (733)	.56 (733)
Prosocial behavior	6.81	3.03	−.58 (488)	−.20 (488)	−.56 (488)	7.34	2.97	−.70 (733)	−.33 (733)	−.71 (733)

Note. According to Cohen (1992), correlations of .1, .3, and .5 may be considered small, medium, and large, respectively. All correlation coefficients were significant to $p < .0001$ with the exception of *a*, which was not significant, and *b*, which was significant to $p < .01$. SRSS-EC = Student Risk Screening Scale for Early Childhood version; SDQ = Strengths and Difficulties Questionnaire.

correlations <.35. We considered removing *lies* due to the item-total correlation below criteria. Given that the item met established criteria for mean, skewness, and kurtosis values in the item-level analyses, the item was retained.

Next, we computed the alpha coefficient for the four items in Factor 2, proposed to measure internalizing behaviors (SRSS-EC I4). The overall alpha was .73, which was lower than the alpha coefficient for the externalizing items as expected. Item-total correlations ranged from a low of .37 for *shy, timid* to a high of .66 for *worried, fearful*, all exceeding the .35 criteria.

Upon looking at the internal consistency of the full scale, the SRSS-EC 11 had an alpha coefficient of .81. This overall alpha coefficient remained about the .80 standard.

Consistent with Lane, Menzies, Oakes, Lambert, et al. (2012) and Lane, Oakes, et al. (2012), decisions regarding item retention were made using empirical and theoretical guidelines as described above. Empirical guidelines included results of item-level descriptive statistics and factor structure. Theoretical considerations also informed decision-making, with attention to retaining items most closely related to core characteristic externalizing and internalizing behaviors noted in the literature. Four questions posed by Hatcher (1994) guided interpretation: (a) Is there a minimum of three items with sufficient loadings on the two factors retained? (b) Do the variables loading on the two factors share conceptual meaning? (c) Do the variables loading on other factors appear to measure distinct constructs? (d) Does the rotated factor pattern yield simple structure? Simple structure refers to two circumstances. First, most items have relatively high factor loadings on one factor and low loading for other factors. Second, most factors have relatively high factor loadings for some items and low factor loadings for other variables. Based on the guidelines, three items were removed: *steals* (a priori), *talks with other children*, and *plays alone*. Then, we examined the validity of the resulting tool: SRSS-EC 11.

Validity

Convergent validity: SDQ. Pearson's product moment correlation coefficients were computed between the SDQ scores (total and subscales) and the following SRSS-EC scores: SRSS-EC E7 (externalizing subscale), the SRSS-EC I4 (internalizing subscale), and SRSS-EC 11 (total score). Total scores ranged from 0 to 21 for the SRSS-EC E7, 0 to 12 for the SRSS-EC I4, and 0 to 33 for the SRSS-EC 11 (see Table 5). The SRSS-EC E7 score was significantly correlated with all SDQ scores to the $p < .0001$ level, with correlation coefficients as follows: total score, $r = .79$; emotional symptoms, $r = .21$; conduct problems, $r = .81$; hyperactivity, $r = .73$; and peer problems, $r = .44$. The SRSS-EC E7 was moderately negatively correlated with the prosocial subscale, $r = -.58$. The SRSS-EC I4 was significantly correlated with SDQ scores: total score, $r = .42$; emotional symptoms, $r = .76$; and peer problems, $r = .35$. As anticipated, correlation coefficients were lower for conduct problems ($r = .08$) and hyperactivity ($r = .13$) subscales. The SRSS-EC I4 was negatively correlated with the prosocial subscale ($r = -.20$). The SRSS-EC 11 was also significantly correlated with all SDQ scores, with correlation coefficients moderately-to-highly positive as follows: total score, $r = .82$; emotional symptoms, $r = .50$; conduct problems, $r = .69$; hyperactivity, $r = .65$; and peer problems, $r = .50$. The SRSS-EC 11 was moderately negatively correlated with the prosocial subscale ($r = -.56$). The SRSS-EC 11 ($r = .82$), relative to the SRSS-EC E7 ($r = .79$) and the SRSS-EC I4 ($r = .42$), had the highest correlation with the SDQ total score. This is likely because the SDQ total score is a combination of externalizing (conduct problems, hyperactivity) and internalizing (emotional symptoms, peer problems). In sum, there is initial evidence to establish convergent validity between SRSS-EC (SRSS-EC E7, SRSS-EC I4, SRSS-EC 11) and SDQ scores.

Table 6. Convergent Validity: SRSS-EC E7, SRSS-EC I4, and SRSS-EC 11 With the ESP Subscale Scores.

	Study 1 (*N* = 489)					Study 2 (*N* = 737)				
			SRSS-EC E7	SRSS-EC I4	SRSS-EC 11			SRSS-EC E7	SRSS-EC I4	SRSS-EC 11
ESP subscales	*M*	*SD*	*r*, *p* (*n*)	*r*, *p* (*n*)	*r*, *p* (*n*)	*M*	*SD*	*r*, *p* (*n*)	*r*, *p* (*n*)	*r*, *p* (*n*)
CEI										
Boys	0.30	0.67	.17, .0452 (136)	.15, .0875 (136)	.23, .0077 (136)	0.21	0.55	.20, .0302 (117)	.30, .0009 (117)	.32, .0004 (117)
Girls	0.23	0.56	.22, .0477 (78)	.24, .0369 (78)	.31, .0050 (78)	0.14	0.46	.05, .6763 (71)	.32, .0071 (71)	.25, .0332 (71)
CFI-A										
Boys	1.45	1.32	.47, <.0001 (157)	−.01, 0.8673 (157)	.43, <.0001 (157)	0.77	1.11	.66, <.0001 (123)	−.02, .8092 (123)	.58, <.0001 (123)
Girls	0.49	0.80	.55, <.0001 (57)	.21, .1119 (57)	.49, .0001 (57)	0.36	0.72	.35, .0042 (64)	.03, .8054 (64)	.29, .0213 (64)
Combined Frequency: Maladaptive Behavior (CFI-M)										
Boys	1.37	1.36	.84, <.0001 (134)	−.13, .1458 (134)	.73, <.0001 (134)	0.91	1.24	.82, <.0001 (116)	−.04, .7057 (116)	.77, <.0001 (116)
Girls	0.88	1.23	.62, <.0001 (78)	.06, .5734 (78)	.56, <.0001 (78)	0.60	1.07	.64, <.0001 (68)	−.03, .7813 (68)	.52, <.0001 (68)
ABS										
Boys	2.09	1.28	.70, <.0001 (79)	.01, .9335 (79)	.61, <.0001 (79)	1.40	1.37	.70, <.0001 (60)	.26, .0421 (60)	.66, <.0001 (60)
Girls	1.76	1.33	.53, .0034 (29)	.27, .1516 (29)	.53, .0031 (29)	0.80	1.22	.75, <.0001 (31)	.42, .0172 (31)	.73, <.0001 (31)
SIS										
Boys	0.63	0.49	−.04, .7826 (57)	.17, .2124 (57)	.03, .7992 (57)	0.53	0.50	.26, .0561 (53)	.26, .0649 (53)	.40, .0029 (53)
Girls	0.51	.051	.37, .0084 (49)	.22, .1297 (49)	.38, .0079 (49)	0.48	.051	.43, .0061 (40)	.48, .0018 (40)	.49, .0014 (40)

Note. According to Cohen's (1992) correlations of .1, .3, and .5 may be considered small, medium, and large, respectively. SRSS-EC = Student Risk Screening Scale for Early Childhood version; ESP = Early Screening Project; CEI = Critical Events Index; CFI-A = Adaptive Behavior Scale; CFI-M = Combined Frequency: Maladaptive Behavior; ABS = Aggressive Behavior Scale; SIS = Social Interaction Scale.
CEI: Boys and Girls—low-risk (0,1), at risk (1 *SD*, 2), high risk (1.5 *SD*, 3), and extreme risk (2 *SD*, 4–14).
CFI-A: Boys—low-risk (28–40), at risk (1 *SD*; 25–27), high risk (1.5 *SD*, 22–24), and extreme risk (2 *SD*; 0–21); Girls—low-risk (30–40), at risk (1 *SD*, 27–29), high risk (1.5 *SD*, 24–26), and extreme risk (2 *SD*, 0–23).
CFI-M: Boys—low-risk (0–19), at risk (1 *SD*; 20–22), high risk (1.5 *SD*, 23–25), and extreme risk (2 *SD*; 26–45); Girls—low-risk (0–19), at risk (1 *SD*, 20–22), high risk (1.5 *SD*, 23–25), and extreme risk (2 *SD*, 26–45).
ABS: Boys—low-risk (0–14), at risk (1 *SD*; 15–16), high risk (1.5 *SD*, 17–18), and extreme risk (2 *SD*; 19–45); Girls—low-risk (0–13), at risk (1 *SD*, 14), high risk (1.5 *SD*, 15), and extreme risk (2 *SD*, 16–45).
SIS: Boys and Girls—low-risk (28–56) and at risk (1 *SD*; 0–27).

Convergent validity: ESP. Spearman's correlation coefficients were computed between the ESP subscale categorical scores and the following SRSS-EC scores: SRSS-EC E7, SRSS-EC I4, and SRSS-EC 11. The SRSS-EC E7 score was significantly correlated with several of the ESP subscale scores. As expected, correlation coefficients were largest for CFI-M and ABS scores (particularly for boys), both of which reflect externalizing behaviors (see Table 6). The SRSS-EC I4 score was significantly correlated with just the CEI of the ESP subscale scores for girls. The SRSS-EC 11 score was significantly correlated with several ESP subscale scores. Relations were largest for CFI-M and ABS scores, suggesting that the SRSS-EC total scores may be more reflective of externalizing than internalizing behaviors. In sum, there is initial evidence to establish convergent validity between the SRSS-EC scores and ESP subscale scores.

Study 2: Method

Participants and Setting

In Study 2, participants included 737 children (425 [57.82%] boys) attending a large, economically diverse public early childhood center in a Midwestern state also classified as Suburban: Large according to locale codes (National Center for Education Statistics, n.d.). Center C was not classified as a Title I school. Each child was rated by their classroom

teacher ($N = 36$) on three screening tools: SRSS-EC, SDQ, and Stages 1 and 2 of the ESP. Most children were White (88.57%; $n = 651$), with 23.78% ($n = 175$) receiving special education services. Students ranged in age from 3.08 to 5.58 years ($M = 4.50$, $SD = 0.61$).

Teachers rated between 5 and 42 students as some teachers taught both morning and afternoon classes. Teachers rated on average 21.06 children ($SD = 9.59$). All teachers were White females ranging in age from 26 to 59 years ($M = 38.26$, $SD = 10.66$). Teachers had between 1 and 31 years of teaching experience as an early childhood teacher ($M = 9.77$, $SD = 7.10$).

Procedures and Measures

Procedures for Study 2 were the same as those specified for Study 1, with two exceptions. At Center C, teachers ($n = 36$) completed the screening packets in their buildings rather than during the consenting meeting. Also, administrators in Center C provided the child demographic data to teachers with the goal of saving teacher time to complete this form.

Statistical Analysis

The analytic plan for Studies 1 and 2 was the same, except for the factor structure. In Study 2, we conducted a CFA. After an initial model is established in an EFA, the next step is to perform a CFA to confirm that the hypothesized model provides a good fit to the data. We examined several fit statistics to determine if the model provides adequate fit for the data as recommended by Holtzman and Vezzu (2011): (a) the chi-square test, which indicates the amount of difference between expected and observed covariance matrices (yet, sensitive to sample size; Joreskog, 1969); (b) the root mean square error of approximation (RMSEA), which is related to residuals in the model, with smaller values indicating better model fit (range = 0–1, <.08 considered acceptable; Browne & Cudeck, 1993); and (c) the comparative fit index (CFI), which examines overall improvement of a given model over an independence model in which observed variables are uncorrelated (range = 0–1, with larger values indicating a better fit, with ≥.95 considered optimal; Hu & Bentler, 1999).

Study 2: Results and Summary

Reliability

Item-level analyses. We again began by examining psychometric properties of the SRSS-EC. Mean scores ranged from 0.07 for *steals* and 0.14 for *lies* (two covert behaviors) to 1.09 for *plays alone*. The *steals* item had two other indicators of concern (skewness and kurtosis).

Factor structure. Findings of one fit index suggested that the model was likely not adequate, yet results of two fit indices were marginal. RMSEA was lower than the <.08 criteria at .061, suggesting an acceptable fit with the two-factor structure. Also, the CFI index flight slightly below the .9 criteria at .89. Despite the model not being an acceptable fit, all parameter estimates were statistically significant from 0, with t statistics all greater than 2.56 (range = 6.82–29.84). Covariance among latent variables (referred to exogenous variables) indicated a t statistic of 12.86, significant at the .01 level. Given the variances are fixed to be 1, the correlations are equal to the covariances. The latent constructs are moderately correlated with an estimate of .47 (standard error = 0.04). In a good-fitting model, most (if not all) parameter estimates are significant as in this case. Thus, the results of this CFA are mixed, indicating that additional inquiry is needed to confirm the factor structure of the SRSS-EC.

Internal consistency. Next, we computed Cronbach's alpha coefficients for the subscale and total scale scores. For SRSS-EC E7, the overall alpha was even higher than in Study 1 with a value of .88 as was the overall alpha for the SRSS-EC 11 with a value of .83. However, the overall alpha was lower for SRSS-EC I4 with a value of .60.

Validity

Convergent validity: SDQ. The SRSS-EC E7 score was significantly correlated with all SDQ scores to the $p < .0001$ level, with a moderately negative correlation with the prosocial subscale, $r = -.70$ (see Table 5). The SRSS-EC I4 was significantly correlated with SDQ scores, with correlation coefficients moderately-to-highly positive as follows: total score, $r = .40$; emotional symptoms, $r = .66$; and peer problems, $r = .35$. As anticipated, correlation coefficients were lower for conduct problems ($r = .15$) and hyperactivity ($r = .21$) subscales. Furthermore, the SRSS-EC I4 was negatively correlated with the prosocial subscale ($r = -.33$). The SRSS-EC 11 was also significantly correlated with all SDQ scores to the $p < .0001$ level, with correlation coefficients moderately-to-highly positive. The SRSS-EC 13 was moderately negatively correlated with the prosocial subscale ($r = -.71$). As in Study 1, the SRSS-EC 11 ($r = .84$), relative to the SRSS-EC E7 ($r = .83$)—although just slightly—and the SRSS-EC I4 ($r = .40$), had the highest correlation coefficient with the SDQ total score. The SRSS-EC 11 ($r = -.71$) also had the highest correlation coefficient with prosocial behavior relative to the SRSS-EC E7 ($r = -.70$) and the SRSS-EC I4 ($r = -.33$). In sum, Study 2 offers additional evidence of convergent validity between SRSS-EC and SDQ scores.

Convergent validity: ESP. Spearman's correlation coefficients were computed between the ESP subscale categorical scores

and the following SRSS-EC scores. In general, results were very similar to those reported in Study 1 (see Table 6). The SRSS-EC E7 score was significantly correlated with the following ESP subscale scores. As with Study 1, correlation coefficients were largest for CFI-M and ABS scores suggesting that SRSS-EC E7, CFI-M, and ABS scores were measuring externalizing behaviors. SRSS-EC I4 scores were significantly correlated with just two ESP subscale scores: CEI for boys and girls, ABS for boys and girls, and SIS for girls. SRSS-EC 11 scores were significantly correlated with all ESP subscale scores. Again, results are similar to Study 1 findings, with the relations largest for CFI-M and ABS scores. In sum, findings offer additional evidence of convergent validity between SRSS-EC and ESP scores.

Discussion

Systematic screening tools are being introduced as central components of tiered systems of support in an effort to accurately detect and subsequently assist students for whom Tier 1 efforts in isolation are insufficient (Lane & Walker, in press). It is encouraging to see the availability of such screenings across the PK–12 continuum for use in identifying students with internalizing and externalizing behaviors, representing both major disorders of childhood. Recognizing that some schools and districts have limited financial resources as well as a strong interest in protecting teachers' time, it is important to have reliable and valid screening tools that are free access and require nominal teacher time to prepare, complete, and score.

To this end, the SRSS has been recently adapted to expand its use to detect internalizing behavior patterns in elementary and middle school settings (Lane, Menzies, Oakes, Lambert, et al., 2012; Lane, Oakes, Carter, et al., 2013; Lane, Oakes, et al., 2012). In this article, we explored the possible utility of a free-access, very brief screening tool modeled after the SRSS-IE, modified to reflect developmentally appropriate behavioral performance patterns of preschool-age students.

In the studies, we offer initial evidence for the reliability and validity of this new tool: the SRSS-EC. Items were representative of salient characteristics of preschool-age students with internalizing and externalizing behaviors as noted in the literature and current screening tools.

Initial Evidence of Reliability

In Study 1, we explored the SRSS-EC using classical test theory by examining item-level data, internal consistency estimates, and factor structure with data collected from two early childhood centers, representing 489 preschoolers. Collective findings of initial descriptive procedures, internal consistency estimates, and EFA suggested that three of the initially introduced items be removed according to theoretical and empirical considerations: *steals, talks with other children*, and *plays alone.*

Findings of an EFA of all 11 items (SRSS-EC 11) indicated a two-factor solution, with items *tantrums (*1), *active, restless* (2), *rejected by peers* (4), *ignores teacher and class rules* (5), *negative attitude* (6), *aggressive behavior* (7), and *lies* (8) loading on Factor 1 (externalizing construct) and items *shy timid* (9), *sad, tearful* (10), *worried, fearful* (11), and *physical complains* (14) loading on Factor 2 (internalizing construct). Results indicated improved simple structure for the SRSS-EC 11 relative to the SRSS-EC 13, which would not have been appropriate given only two items loaded on Factor 3 (which may have been measuring social engagement: *talks with other children* and *plays alone*).

Empirical guidelines and theoretical considerations were used to drive decision-making. The answers to Hatcher's (1994) questions for interpreting findings also supported the two-factor solution. First, a minimum of three items with significant loadings on each of the two factors were retained. Second, items loading on respective factors shared conceptual meaning, suggesting that SRSS-EC E7 items were measuring the underlying construct of externalizing behavior and SRSS-EC I4 items were measuring the underlying construct of internalizing behaviors. Third, items loading on Factors 1 and 2 appeared to assess different constructs. Fourth, the rotated factor pattern demonstrated simple structure. Collectively, results supported the decision in Study 1 to remove items: *steals, talks with other children*, and *plays alone.*

In Study 2, findings paralleled Study 1 results, yet the alpha coefficient for the internalizing subscale decreased to .60 (an unacceptable value). Results of the CFA were mixed with only one (RMSEA) criterion clearly suggesting that the model was adequate. Yet, parameter estimates were statistically significant from 0 (suggesting a good-fitting model), providing support for this two-factor structure. Given the mixed results of Study 2 pertaining to the reliability of the SRSS-EC, we encourage cautious optimism at this point. We suggest that other studies be conducted to establish the reliability of this two-factor screening tool for use at the preschool level before recommending the use of the screening tool outside of controlled studies.

Initial Evidence of Convergent Validity

The next objective was to examine convergent validity of the SRSS-EC with other established screening tools: the SDQ and the ESP. Similar to the initial results reported by Lane, Menzies, Oakes, Lambert, et al. (2012) and Lane, Oakes, et al. (2012) comparing the SRSS-IE scores with SDQ and SSBD scores, preliminary findings of Study 1 indicated convergent validity between the SRSS-EC scores with the SDQ and ESP scores with the preschool sample.

Convergent validity: SDQ. In terms of the SDQ, convergent validity between SRSS-EC E8 and SDQ scores was consistent with those reported by Lane, Oakes, et al. (2012) and Lane, Menzies, Oakes, Lambert, et al. (2012) at the elementary level between SRSS-E7 scores and SDQ subscale scores, with the current results of Study 1 yielding correlations as follows: the total ($r = .79$), conduct problems ($r = .81$), hyperactivity ($r = .73$), and prosocial ($r = -.58$).

Similarly, the SRSS-EC I5 was also consistent with findings reported by Lane and colleagues at the elementary level. Moreover, Study 1 indicated statistically significant, positive correlation between SRSS-EC I4 scores and SDQ emotional symptoms ($r = .76$), and peer problems ($r = .35$) subscales (both characteristic behaviors of internalizing patterns) and low-magnitude positive correlations with SDQ conduct problems ($r = .08$) and hyperactivity ($r = .13$) subscales (more consistent with characteristics of externalizing behaviors).

Finally, the correlation coefficients between the SRSS-EC 11 and all SDQ scores were also statistically significant, yielding a particularly large relation between SRSS-EC 11 and the SDQ total scores ($r = .82$), suggesting that the full scale scores for both measures are capturing internalizing and externalizing dimensions. Likewise, the large, negative correlation between the SRSS-EC 11 and the SDQ prosocial dimension ($r = -.56$) suggesting that (a) the former measure is assessing the opposite of prosocial behavior, namely, antisocial behavior and (b) students with internalizing as well as externalizing behavior patterns struggle socially (H. M. Walker, Ramsey, & Gresham, 2004).

Results from Study 2 confirmed these same findings, supporting the generalizability of Study 1 outcomes. Collectively, findings from Studies 1 and 2 suggest that the SRSS-EC E7, the SRSS-EC I4, and the SRSS-EC 11 scores have convergent validity with the SDQ scores, indicating that measures are accurate in detecting both major disorders of childhood.

Convergent validity: ESP. In terms of the ESP, Spearman's correlation coefficients offer initial evidence of convergent validity between SRSS-EC E7 and ESP subscale scores, with the largest relation established between maladaptive scores and aggressive behavior scores (particularly for boys with respective correlation coefficients of .84 and .70). Both these subscales (CFI-M and ABS) are the most proximal characteristics of externalizing behavior.

The SRSS-EC I4 score was most highly correlated with the CEI scored with a correlation coefficient of .24 for girls. It may be that teachers were only aware of the more serious behavioral markers or that the ESP subscale may have been more reflective of externalizing behaviors. Yet, given the lower internal consistency estimates, it is more likely that additional items will need to be added to the SRSS-EC I4 to better measure internalizing behaviors.

Finally, correlation coefficients between the SRSS-EC 11 and all ESP scores were also statistically significant with the exception of SIS scores for boys. The magnitude of the relations was most pronounced for maladaptive and aggressive behavior subscale scores, which were contrary to our expected outcomes. Thus, while the SRSS-EC is measuring aspects of internalizing and externalizing behaviors, the SRSS-EC total scores may be more indicative of externalizing behaviors, given that the magnitude of the relations was highest for maladaptive and aggressive behavior scores and lowest for social interaction scores.

Results from Study 2 were similar to Study 1's outcomes, offering evidence of generalizability of the findings. Moreover, Study 2 results also suggested that SRSS-EC E7 and SRSS-EC 11 scores had the largest correlations with maladaptive and aggressive behavior scores, whereas SRSS-EC I4 scores had the largest correlations with social interaction scores. Collectively, findings from Studies 1 and 2 suggest that the SRSS-EC E7, the SRSS-EC I4, and the SRSS-EC 11 scores have convergent validity with ESP subscale scores, indicating that the measures are accurate in detecting both major disorders of childhood in a reliable and valid manner—yet, less accurate on internalizing dimension.

We are encouraged by results of these first two psychometric studies of the SRSS-EC, which offer preliminary evidence to support the reliability and validity of the SRSS-EC for use at the preschool level. Yet, we do not want to overstate the findings, and we emphasize that it is critical for other studies to be conducted to establish generalizability of reported results. As such, we interpret findings of the studies presented with attention to the following limitations.

Limitations and Directions for Future Research

While findings from Studies 1 and 2 were highly consistent with the convergent validity findings reported by Lane, Oakes, et al. (2012) and Lane, Menzies, Oakes, Lambert, et al. (2012) regarding the SRSS-IE, we emphasize the need for additional inquiry at the preschool level to address concerns regarding generalizability. Samples in Studies 1 and 2 were from one state. While participants were culturally and ethnically diverse, additional studies are needed to examine the generalizability of findings with other students and in other locales (Lane, Oakes, et al., 2012).

Second, data reported in Studies 1 and 2 reflect one time point. Moving forward, it is important for other research teams to examine the reliability and validity of the SRSS-EC over time within (e.g., fall, winter, and spring) and across (e.g., 2014–2015 to 2015–2016) academic years. Other psychometric studies of screening tools have suggested that SRSS scores (the initially developed tool) become more accurate in predicting students' academic and behavioral outcomes when completed later in the school year and at the high school level, later in the school day (see Lane, Oakes,

Ennis, et al., 2013). Logically, it is reasonable to expect that as teachers have more time with students, teachers become more familiar with student behavior as the year progresses.

Third, and related to the second limitation, this study did not examine predictive validity of the SRSS-EC. The validity evidence offered preliminary evidence of convergent validity with two established screening tools: the SDQ and ESP. While this is a good starting point, the next question to ask is, "Do SRSS-EC scores predict important behavioral, social, and early academic skills for preschool-age students over 1 year and as they transition to kindergarten?" As such we recommend additional inquiry to build on the initial evidence offered by these first two studies.

Fourth, the order in which teachers completed the three screening tools was not counterbalanced which could have controlled for potential order effects and we did not assess teachers' views of the screening tools (social validity). Given that teachers were using two screening tools for which they had no experience (SRSS-EC, SDQ) and our goal was to provide clear instructions and be available to answer questions, we did not counterbalance. Also, given that a social validity measure would only be able to reflect opinions about completion rather than how the data informed intervention efforts and because we wanted to limit time required for participation, we did not assess social validity. We encourage others to explore these important issues.

Fifth, while the current study is an adequate initial psychometric study of the SRSS-EC (consistent with the initial study conducted for the SRSS-IE), there is ample methodological work to be conducted. Ideally, subsequent studies would include a large enough teacher sample to address the nested nature of the data, meaning that the potential consequences of dependencies of a single teacher rating multiple students, and in some instances rating two classrooms of students which was not addressed in this study. While the precise consequences of nesting are not known, it is possible student-level screening scores may reflect rater as well as student characteristics. This may inflate internal consistency reliability estimates, which is why subsequent studies are needed to explore this consideration (Elliott & Gresham, 2007). Other studies are needed to establish cut scores to determine which scores for the SRSS-EC E7, SRSS-EC I4, and SRSS-EC 11 suggest low, moderate, and high risk status as this information is highly useful for schools in connecting students to subsequent supports (e.g., Oakes, Mathur, & Lane, 2010).

Finally, the SRSS-EC is a teacher-completed screening tool for the early detection of students who may need additional instruction, intervention, and practice with social and behavioral skills sets within the education program. Given the importance of family involvement and recognition of the family as the student's primary teacher—particularly at the preschool level—future studies should address the use of the SRSS-EC for family-centered planning (Division of Early Childhood of the Council for Exceptional Children, 2014). Ideally, studies could also explore the relation between SRSS-EC and ASQ:SE, with attention to issues of utility and validity. It encourages others to examine how these tools each contribute to offer additional insight into students' skill sets (e.g., self-regulation; McClelland & Cameron, 2010) and additional perspectives (i.e., parents, guardians, caregivers; Neisworth & Bagnato, 2005).

Summary

Despite the limitations noted, these first two studies offer evidence to suggest the SRSS-EC 11 as a two-factor measure, yielding two subscales scores: the SRSS-EC E7, which appears to be measuring the construct of externalizing behaviors exhibited by preschool students, and the SRSS-EC I4, which appears to be measuring the construct of internalizing behaviors (Benson, 1998). Results of convergent validity analyses indicate that the SRSS-EC scores have convergent validity with established screening tools: SDQ and ESP. While the simplicity of this free-access and very brief screening tool may hold promise for preschool educators interested in using systematic screening tools to accurately detect young students who may need subsequent support in the form of Tier 2 and 3 opportunities, we caution that this tool is not yet ready for such use until subsequent studies are conducted exploring predictive validity. Again, while we are encouraged by the findings reported, we reiterate the importance of establishing generalizability before using this tool to work within tiered systems of support at the preschool level.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YRS, & TRF profiles*. Burlington: Department of Psychiatry, University of Vermont.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, *1*, 10–17.

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

Cattell, R. B. (1966). The scree plot for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cooper, J. L., & Vick, J. (2009). *Promoting social-emotional well-being in early intervention services: A fifty-state view*. New York, NY: National Center for Children in Poverty.

Coplan, R. J., & Rubin, K. H. (1998). Exploring and assessing non-social play in the preschool: The development and validation of the Preschool Pay Behavior Scale. *Social Development*, *7*, 72–91.

Costello, E. J., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry*, *47*, 1263–1271.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*, 391–418.

Division of Early Childhood of the Council for Exceptional Children. (2014). *Recommended practices in early intervention and early childhood special education*. Los Angeles, CA: Division of Early Childhood.

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.

Dunlap, G., Sailor, W., Horner, R., & Sugai, G. (2009). Overview and history of positive behavior support. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 3–16). New York, NY: Springer. doi:10.1007/978-0-387-09632-2_1

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*, 313–337.

Elliott, S. N., & Gresham, F. M. (2007). *Social skills improvement system: Performance screening guides*. Bloomington, MN: Pearson Assessments.

Feeney-Kettler, K. A., Kratochwill, T. R., Kaiser, A., Hemmeter, M. L. P., & Kettler, R. J. (2010). Screening young children's risk for mental health problems: A review of four measures. *Assessment for Effective Intervention*, *35*, 218–230. doi:10.1177/1534508410380557

Feil, E. G., Severson, H. H., & Walker, H. M. (1995). Identification of critical factors in the assessment of preschool behavior problems. *Education & Treatment of Children*, *18*, 261–271.

Forness, S. R., Freeman, S. F. N., Paparella, T., Kauffman, J. M., & Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *20*, 4–18. doi:10.1177/1063426611401624

Fox, L., Carta, C., Strain, P., Dunlap, G., & Hemmeter, M. L. (2010). Response to intervention and the Pyramid Model. *Infants & Young Children*, *23*, 3–13.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117–135.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 1337–1345.

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

Hatcher, L. (1994). *A step-by-step approach to using SAS system for factor analysis and structural equation modeling*. Cary, NC: The SAS Institute.

Hemmeter, M. L., & Conroy, M. (2012). Supporting the social competence of young children with challenging behavior in the context of the Teaching Pyramid model: Research based practices and implementation in early childhood settings. In R. Pianta, L. Justice, S. Barnett, & S. Sheridan (Eds.), *The handbook of early education* (pp. 416–434). New York, NY: Guilford Press.

Holtzman, S., & Vezzu, S. (2011). *Confirmatory factor analysis and structural equation modeling of noncognitive assessments using PROC CALIs*. Princeton, NJ: Educational Testing Service.

Hopkins, W. G. (2002). A scale of magnitudes for the effect statistics. In *A new view of statistics*. Retrieved from http://www.sportsci.org/resource/stats/index.html

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *32*, 179–185.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Joreskog, K. G. (1969). A general approach to confirmatory factor analysis. *Psychometrika*, *34*, 183–202.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 behavior and emotional screening system* (BASC-2 BESS). San Antonio, TX: Pearson.

Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (Eds.). (2014). *Universal screening in educational settings: Evidence-based decision making for schools*. Washington, DC: American Psychological Association.

Lane, K. L., Kalberg, J. R., & Menzies, H. M. (2009). *Developing schoolwide programs to prevent and manage problem behaviors: A step-by-step approach*. New York, NY: Guilford Press.

Lane, K. L., & Menzies, H. M. (in press). Classroom management in inclusive settings. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook of classroom management* (2nd ed.). Taylor & Francis.

Lane, K. L., Menzies, H. M., Oakes, W. P., & Kalberg, J. R. (2012). *Systematic screenings of behavior to support instruction: From preschool to high school*. New York, NY: Guilford Press.

Lane, K. L., Menzies, H. M., Oakes, W. P., Lambert, W., Cox, M. L., & Hankins, K. (2012). A validation of the Student Risk Screening Scale for internalizing and externalizing behaviors: Patterns in rural and urban elementary schools. *Behavioral Disorders*, *37*, 244–270.

Lane, K. L., Oakes, W. P., Carter, E. W., Lambert, W., & Jenkins, A. (2013). Initial evidence for the reliability and validity of the Student Risk Screening Scale for Internalizing and Externalizing Behaviors at the middle school level. *Assessment for Effective Intervention*, *39*, 24–38. doi:10.1177/1534508413489336

Lane, K. L., Oakes, W. P., Ennis, R. P., Cox, M. L., Schatschneider, C., & Lambert, W. (2013). Additional evidence for the reliability and validity of the Student Risk Screening Scale at the high school level: A replication and extension. *Journal of Emotional and Behavioral Disorders*, *21*, 97–115.

Lane, K. L., Oakes, W. P., Harris, P. J., Menzies, H. M., Cox, M. L., & Lambert, W. (2012). Initial evidence for the reliability and validity of the Student Risk Screening Scale for internalizing and externalizing behaviors at the elementary level. *Behavioral Disorders*, *37*, 99–122.

Lane, K. L., & Walker, H. M. (in press). The connection between assessment and intervention: How does screening lead to better interventions? In B. Bateman, M. Tankersley, & J. Lloyd (Eds.), *Issues in special education*.

McClelland, M. M., & Cameron, C. E. (2010). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*, 136–142.

National Center for Education Statistics. (n.d.). *Identification of rural locales*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/ccd/rural_locales.asp

Neisworth, J. T., & Bagnato, S. J. (2005). DEC recommended practices: Assessment. In S. Sandall, M. L. Hemmeter, B. J. Smith, & M. E. McLean (Eds.), *DEC recommended practice: A comprehensive guide for practical application in early intervention/early childhood special education* (pp. 45–69). Longmont, CO: Sopris West.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). St. Louis, MO: McGraw-Hill.

Oakes, W. P., Mathur, S. R., & Lane, K. L. (2010). Reading interventions for students with challenging behavior: A focus on fluency. *Behavioral Disorders*, *35*, 120–139.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, & Computers*, *32*, 396–402.

Squires, J., Bricker, D., & Twombly, E. (2002). *Ages & Stages Questionnaires*®: *Social-Emotional (ASQ:SE)*. Baltimore, MD: Paul H. Brookes.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327.

Walker, H. M., Ramsey, E, & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (SSBD): User's guide and technical manual*. Longmont, CO: Sopris West.

Walker, H. M., Severson, H. H., & Feil, E. G. (1995). *The early screening project: A proven child-find process*. Longmont, CO: Sopris West.

Walker, L. S., Beck, J. E., Garber, J., & Lambert, W. (2009). Children's Somatization Inventory: Psychometric properties of the revised form (CSI-24). *Journal of Pediatric Psychology*, *34*, 430–440.

Wichstrom, L., Berg-Nielsen, T. S., Angold, A., Eagger, H. L., Solheim, E., & Sveen, T. H. (2012). Prevalence of psychiatric disorders in preschoolers. *The Journal of Child Psychology and Psychiatry*, *53*, 695–705. doi:10.1111/j.1469-7610.2011.02514.x