Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2019, Vol. 27(4) 221–234

DOI: 10.1177/1063426618795443
jebd.sagepub.com
SAGE

Predictive Validity of Student Risk Screening Scale—Internalizing and Externalizing (SRSS-IE) Scores in Elementary Schools

Kathleen Lynne Lane[1], Wendy Peia Oakes[2], Emily D. Cantwell[1], Eric A. Common[3], David J. Royer[4], Melinda M. Leko[5], Christopher Schatschneider[6], Holly Mariah Menzies[7], Mark Matthew Buckman[1], and Grant Edmund Allen[1]

Abstract

In this article, we examined predictive validity of *Student Risk Screening Scale—Internalizing and Externalizing* (SRSS-IE) scores for use with elementary-age students (N = 4,465) from 14 elementary schools. Results indicated elementary school students with high levels of risk according to fall SRSS-IE scores—especially those with externalizing behaviors—were more likely to have lower oral reading fluency scores, lower Measures of Academic Progress (MAP) reading scores, more nurse visits, and more days spent in in-school suspension compared with students at low risk for externalizing or internalizing behaviors. Educational implications, limitations, and future directions are presented.

Keywords

systematic screening, externalizing, internalizing, elementary, predictive validity

Throughout the United States, federal, state, and local educational leaders have placed a high priority on developing integrated tiered systems of support such as the comprehensive, integrated, three-tiered (Ci3T) models of prevention to meet students' academic, behavioral, and social needs (Lane, Kalberg, & Menzies, 2009; McIntosh & Goodman, 2016; Yudin, 2014). Such tiered systems offer a cascade of evidence-based strategies, practices, and programs for students at each level of prevention: primary (Tier 1) for all, secondary (Tier 2) for some, and tertiary (Tier 3) for few (Cook & Tankersley, 2013). The Ci3T model creates a structure for preventing the development of learning and behavior challenges from arising and responding effectively and efficiently when such challenges do arise (Lane, Oakes, Cantwell, & Royer, 2016). A keystone feature of tiered systems is data-informed decision making, with academic and behavior systematic screening data used in tandem to determine how to assist students for whom primary prevention efforts—even when implemented with integrity—are insufficient to meet students' multiple needs (Oakes, Lane, Cox, & Messenger, 2014).

These models may hold particular benefits for students with emotional and behavioral disorders (EBD), a large and diverse group of students who struggle with externalizing (e.g., aggressive) and internalizing (e.g., anxious) behaviors. Externalizing behaviors often disrupt the learning environment by impeding instructional processes creating challenges not only for the student struggling with externalizing behaviors but also for his or her peers and teachers. While internalizing behaviors are often more covert and less apt to negatively impact the learning environment, they are no less serious as they adversely affect interpersonal relationships and academic performance (Bradshaw, Buckley, & Ialongo, 2008). Teachers consistently report managing challenging behaviors as one of the biggest factors impeding effective teaching (New Teacher Project, 2013). Clearly, this is no small challenge.

[1]The University of Kansas, Lawrence, USA
[2]Arizona State University, Tempe, USA
[3]University of Michigan–Flint, USA
[4]University of Hawai'i at Mānoa, Honolulu, USA
[5]University of Wisconsin–Madison, USA
[6]Florida State University, Tallahassee, USA
[7]California State University, Los Angeles, USA

Corresponding Author:
Kathleen Lynne Lane, Department of Special Education, School of Education, The University of Kansas, 1122 West Campus Road, JRP Room 542, Lawrence, KS 66045, USA.
Email: Kathleen.Lane@ku.edu

Forness, Freeman, Paparella, Kauffman, and Walker (2012) report 20% of school-age children and youth demonstrate mild-to-severe EBD. With less than 1% of students typically qualifying for special education services under the category of emotional disturbance as defined in the Individuals With Disabilities Education Improvement Act (2004), this leaves the general education community largely responsible for meeting the multiple needs of student with and at risk for EBD. The Ci3T model may provide the ideal context to meet this formidable charge. It is important to explore feasible and effective solutions to support all students—particularly those with EBD given long-term negative consequences characteristic of this difficult-to-teach population: disengagement, school failure, school dropout, impaired personal relationships, and increased need for mental health supports (Maggin, Wehby, Farmer, & Brooks, 2016).

Fortunately, the Institute of Education Sciences (IES) recognizes potential benefits of tiered systems as a mechanism that "provides academic, social, emotional, and behavioral support for all students, and provides resources and supports that teachers and other school personnel need to support" students with and at risk for learning and behavioral challenges in authentic educational settings (US Department of Education, IES, 2017, p. 17). In the recent request for application (RFA) from IES, research considerations were offered across all topic areas: (a) inquiry on comprehensive, integrated frameworks and (b) research to develop and evaluate adaptive interventions, including individually tailored interventions to assist students with intensive intervention needs. This is but one illustration of how federal agencies such as the U.S. Department of Education (USDOE) have prioritized this type of work, with systematic screening a central feature needed to facilitate inquiry in both these key objectives.

Given the importance of using systematic screening data in conjunction with other data collected on all students as part of regular school practices (e.g., academic assessments, attendance), it is essential for schools to have access to reliable, valid, and feasible screening tools (Lane, Oakes, Ennis, & Royer, 2015). Several behavior screening tools have been developed and refined to accurately detect students with and at risk for EBD (Lane & Walker, 2015). Examples of such tools include Behavior Assessment System for Children 3rd Edition: Behavioral & Emotional Screening System (BASC-3: BESS; Kamphaus & Reynolds, 2015); Social, Academic, and Emotional Behavior Risk Screener© (SAEBRS; Kilgus, Chafouleas, Riley-Tillman, & von der Embse, 2013); Social Skills Improvement System—Performance Screening Guide (SSiS-PSG; Elliott & Gresham, 2008); Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001); Student Risk Screening Scale (SRSS; Drummond, 1994); Student Risk Screening Scale—Internalizing and Externalizing (SRSS-IE; Drummond, 1994; Lane & Menzies, 2009); and Systematic Screening for Behavior Disorders (SSBD; Walker, Severson, & Feil, 2014). Now, the pivotal question facing many district- and school-site leaders is, "Which screening tool should we adopt?" (Lane, Oakes, Ennis, & Royer, 2015).

The selection of a behavior screening tool is an important one, guided by a number of considerations: facets of behavior challenges to be detected (externalizing and/or internalizing), school or grade levels of interest, informant (teacher, parent, and/or student), administration and scoring time, associated costs (purchase price, ongoing costs), and availability of intervention resources (Lane, Menzies, Oakes, & Kalberg, 2012). In addition to logistical considerations, an overarching concern is being certain the screening tool minimizes false negatives (overlooking a student who does have the challenge of interest), and, although less of a priority, minimizes false positive (indicating a student has the challenge of interest when, in fact, they do not). Ideally, educational leaders' decision-making processes would not be driven primarily by monetary considerations. Yet, in light of current fiscal uncertainty of educational funding, monetary concerns are a pragmatic consideration. For some schools, free-access screening tools such as the SDQ, SRSS, and SRSS-IE may be the only realistic options (Lane et al., 2018).

Given these realities, it is imperative for the research community to explore psychometric proprieties of *all* screening tools—commercially available and free access. This inquiry is not conducted to "prove" any one screening tool is the best option but to offer the practitioner community the full scope of information necessary to inform decision-making processes when selecting a screening tool to detect students with and at risk for social, emotional, and behavioral challenges. The National Center on Intensive Intervention (NCII) established the Behavior Screening Technical Review Committee (TRC) to address this charge. In partnership with the Academic Screening and Progress Monitoring TRC groups, the following definition of screening was established: "a process using tools with convincing evidence of classification accuracy, reliability, and validity to identify students who may require intensive intervention efforts to meet their academic, social, emotional and/or behavioral needs" (NCII, 2017).

Responding to (a) the need for rigorous inquiry regarding the classification, reliability, and validity of existing tools and (b) the fact many school systems may need to move forward with free access screening tools due to fiscal challenges, we conducted this predictive validity study of SRSS-IE scores in predicting important educational outcomes. We begin by briefly examining psychometric properties of the SRSS developed by Drummond (1994), followed by a discussion of research conducted on the SRSS-IE—an expanded tool designed to broaden the scope of the SRSS to detect students with internalizing and externalizing behaviors.

Psychometric Properties of the SRSS and SRSS-IE

Several psychometric studies of SRSS scores have examined the utility of this seven-item universal screening tool initially designed to detect students with antisocial tendencies. At the elementary level, teachers independently rate each student in their homeroom class using a 4-point Likert-type scale (*never* = 0, *occasionally* = 1, *sometimes* = 2, *frequently* = 3). Items include (a) steal; (b) lie, cheat, sneak; (c) behavior problem; (d) peer rejection; (e) low academic achievement; (f) negative attitude; and (g) aggressive behavior. A composite score is created by summing item-level data (range = 0 to 21), with scores used to place students into one of three categories: low (0–3), moderate (4–8), or high (9–21) risk (Drummond, 1994). Studies offer evidence of score reliability and validity at the elementary level as evidenced by strong internal consistency and additional evidence that fall SRSS scores predicted year-end office discipline referral (ODR) rates and spring oral reading fluency (ORF) scores (Menzies & Lane, 2012; Oakes et al., 2010). In addition, studies offered evidence of convergent validity between SRSS scores and SSBD (Lane, Kalberg, Lambert, Crnobori, & Bruhn, 2010; Lane, Little, et al., 2009) and SSiS-PSG scores (Lane, Richards-Tutor, Oakes, & Connor, 2014). Both the SSBD and SSiS-PSG are established, easy-to-use, commercially available behavior screening tools, with the SSiS offering additional tools such as behavior ratings scales and intervention guides. Noting potential benefits of the feasibility of the SRSS that requires about 15 min to screen an entire homeroom class and the initial evidence of classification accuracy, reliability, and validity, Lane and Menzies (2009) modified the SRSS to add items expanding the scope of the tool to detect students with internalizing behaviors. This new tool was called the SRSS-IE.

Lane, Oakes, et al. (2012) first examined the psychometric properties of the SRSS-IE with a sample 2,460 elementary students from California and Arizona. Five of the initially proposed items to detect internalizing behaviors were retained. These internalizing items included (a) emotionally flat; (b) shy, withdrawn; (c) sad, depressed; (d) anxious; and (e) lonely. The SRSS-IE with all 12 items is rated using the same Likert-type scale introduced by Drummond (1994). In addition to offering initial evidence of reliability, results offered initial evidence of the convergent validity between SRSS-IE scores and SSBD and SDQ scores. Lane, Menzies, Oakes, Lambert, et al. (2012) conducted two replication studies, examining psychometric properties of SRSS-IE scores with students in rural (N = 982) and urban (N = 1,079) districts. Results provided additional evidence of reliability, with the same five items retained and additional evidence of convergent validity between SRSS-IE and SSBD scores. Collectively, results supported the utility of SRSS-IE in detecting students with externalizing (SRSS-E7) and internalizing (SRSS-I5) behaviors in a similar fashion to the SSBD.

To further explore convergent validity, Lane, Oakes, Common, et al. (2015) conducted a convergent validity study comparing SRSS-IE and SSiS-PSG scores with a sample of 458 K–5 students from one school in a southeastern state. Correlation analyses indicated statistically significant inverse relations between SRSS-IE (SRSS-E7 and SRSS-I5 subscale scores and the total score) and SSiS-PSG subscale scores. Receiver operating characteristic curve (ROC) analyses comparing scores from students with significant difficulty (highest level of risk) to those making adequate progress (typical performance) indicated SRSS-IE scores were comparable with SSiS-PSG in detecting Prosocial Behavior (area under the curve [AUC] = .972) and Motivation to Learn (AUC = .904). As expected, SRSS-IE scores were less accurate than SSiS-PSG scores in detecting academic risk as that is not the intent of the SRSS-IE (Math Skills AUC = .817; Reading Skills AUC = .805). Lane, Oakes, Ennis, and Royer (2015) conducted a replication study, with results offering comparable findings with a larger sample of 1,680 K–6 students from three schools in a northeastern state.

Collectively, these studies offered evidence of reliability and validity of the SRSS-IE scores, with comparable accuracy to the SSBD, SDQ, and SSiS-PSG scores in detecting students with externalizing and internalizing behaviors. A next important step in this line of inquiry is to explore predictive validity of SRSS-IE scores. Predictive validity refers to the degree to which a score on a scale (e.g., SRSS-E7 or SRSS-I5) predicts scores on a given criterion measure (e.g., number of in-school suspensions or the number of nurse visits, each indicating time away from instruction). As mentioned, fall SRSS (now called SRSS-E7) scores predict a range of academic and behavioral outcomes (e.g., ORF scores, ODR rates). We now intend to replicate predictive validity studies of fall SRSS-E7 scores and explore predictive validity of fall SRSS-I5 scores.

Purpose

In this study, we provide initial evidence to support the utility of SRSS-IE scores for use with elementary students and explore predictive validity of the original SRSS scores. We examined predictive validity of fall SRSS-IE scores by analyzing the degree to which K–5 students with low, moderate, and high risk for externalizing and internalizing behaviors could be differentiated on behavioral and academic characteristics according to extant schoolwide data. We conducted this study to replicate previous inquiry establishing predictive validity of SRSS-E7 scores measuring externalizing behaviors (Menzies & Lane, 2012; Oakes et al., 2010) and examine predictive validity of SRSS-I5

Table 1. Student Characteristics.

Variable/level	*N* = 4,465
Gender % (*n*)	
Male	52.86 (2,360)
Female	47.14 (2,105)
Grade % (*n*)	
Kindergarten	18.23 (814)
First	15.99 (714)
Second	16.55 (739)
Third	17.78 (794)
Fourth	17.18 (767)
Fifth	14.27 (637)
Ethnicity/race % (*n*)	
Hispanic	8.47 (378)
White	72.81 (3,251)
Black	6.61 (295)
Asian/Pacific Islander	4.55 (203)
Native American/Native Alaskan	4.32 (193)
Declined	0.31 (14)
Mixed races	11.40 (509)
Special education % (*n*)	16.84 (752)
Emotional disturbance	31
Intellectual disability	12
Speech language delays	178
Specific learning disabilities	182
Autism spectrum disorder	51
Other health impaired	42
Developmental delay	138
In special education, primary label not reported	12

Note. *N* represents all students enrolled over the course of the academic year. Data are reported for ethnicity (i.e., Hispanic) and race for students. Special education eligibility reported for categories with 10 or more students assigned.

scores measuring internalizing behaviors as applied at the elementary level. We examined ORF as measured by AIMSweb scores, Measures of Academic Progress (MAP), number of nurse visits (as frequent visits could signal a range of concerns), and in-school suspensions. We hypothesized SRSS-E7 and SRSS-I5 scores would be more reflective of behavioral rather than academic outcomes given the former sets of variables are more indicative of constructs measured using the SRSS-IE behavior screening tool (Lane, Oakes, Ennis, & Royer, 2015; Lane, Richards-Tutor, et al., 2014).

Method

Participants and Setting

Participants were 4,465 kindergarten through fifth-grade students (2,360 males) attending one of 14 Midwest elementary schools rated by their homeroom teachers ($n = 219$) on the SRSS-IE. Students were predominantly White (72.81%, $n = 3,251$), with approximately 16.84% receiving special education services (see Table 1). Economic disadvantage rates varied (10.4%–68.7%), with the majority being Title 1 eligible (see Table 2). Schools participating in this study were in the first year of a researcher–practitioner partnership grant funded by IES focused in the implementation and evaluation of Ci3T models of prevention.

Procedures

Ci3T Leadership Teams consisting of the principal, two general education teachers, one special education teacher, an individual with expertise in school-based interventions (e.g., instructional coach, social worker, school psychologist, or behavior specialist), a parent, and a student attended a year-long training series led by university partners to develop a Ci3T model of prevention. As part of the a Ci3T professional learning series, Ci3T Leadership Teams from each of 14 elementary schools reviewed current psychometric evidence on existing screening tools and listed the top three behavior screening tools of interest. District leaders collaborated with Ci3T Trainers to compile these lists, obtain additional information regarding these tools (e.g., procedures for administering, scoring, and interpreting; cost; personnel time), and present information acquired to their district principal leadership team. The district principal leadership team selected the SRSS-IE as this free-access tool took limited teacher time to complete, had established psychometric evidence (e.g., convergent validity with other screening tools, predicted important school outcomes, strong internal consistency), and could be built and maintained at no charge in their existing district database management system.

According to the district assessment schedule, homeroom teachers independently completed the SRSS-IE three times per year: fall (4–6 weeks after the year began), winter (prior to winter break), and spring (before year end). Data from systematic screenings were used by the Ci3T District Leadership Team and Ci3T Leadership Teams from each school to (a) examine overall level of risk schoolwide; (b) inform the use of low-intensity, teacher-delivered supports to increase engagement and decrease disruption; and (c) connect students with Tier 2 and Tier 3 supports as needed (Lane, Kalberg, & Menzies, 2009). Prior to completing the SRSS-IE, teachers received information from their school's Ci3T Leadership Team on the rationale and logistics for completing screening tools. Professional learning opportunities were available for faculty and staff through a range of avenues including afterschool presentations by Ci3T Trainers, districtwide presentations, on-demand resources (e.g., YouTube videos), and practice guides (Lane, Carter, Jenkins, Magill, & Germer, 2015).

Table 2. School Characteristics.

Variable	School						
	ES 1	ES 2	ES 3	ES 4	ES 5	ES 6	ES 7
Enrollment *N*[a]	310	278	522	350	412	512	212
Teachers completing screeners	14	11	23	17	13	23	10
Attendance rate[a] %	95.6	96.2	96.0	96.2	95.1	96.4	94.2
State assessment[a] % (ELA/M)	63.7/49.7	63.9/57.1	56.2/45.6	65.9/58.6	33.3/25.5	71.8/68.0	59.6/51.1
Title 1 school[b]	Yes	Yes	No	Yes	Yes	No	Yes
Economic disadvantaged[a] %	47.1	43.9	29.1	60.0	68.7	10.4	65.6
Students with disabilities[a] %	13.2	11.5	10.0	9.4	40.7	6.3	11.3
	ES 8	ES 9	ES 10	ES 11	ES 12	ES 13	ES 14
Enrollment[a]	226	353	434	383	490	285	229
Teachers completing screeners	10	17	20	18	19	13	11
Attendance rate[a] %	97.0	96.0	96.7	94.7	95.6	96.1	97.0
State assessment[a] % (ELA/M)	58.7/47.6	44.2/36.6	68.1/62.4	36.7/48.1	49.6/51.3	57.0/57.0	43.4/47.8
Title 1 eligible[b]	Yes	Yes	No	Yes	Yes	Yes	Yes
Economic disadvantaged[a] %	63.3	46.2	21.2	60.6	47.8	46.0	51.1
Students with disabilities[a] %	15.5	16.4	8.5	11.2	10.6	13.3	11.8

Note. ES = elementary school; State assessment = percentage reported for students scoring in Levels 3 (at expectations) and 4 (above expectations); ELA = English language arts; M = math; Locale = City Small for all schools.
[a]State school report card data 2015–2016. [b]National Center for Education Statistics, Common Core Data 2014–2015.

The 14 elementary schools, with leadership from each site's Ci3T Leadership Team, began systematic screening for students' academic and behavior performance using AIMSweb and the SRSS-IE during the 2014–2015 academic year. Data presented in this study are from the 2015–2016 academic year, the elementary schools' second year of implementing Ci3T, and the first year the Ci3T District Leadership Team implemented the SRSS-IE districtwide (see Lane et al., 2018 for results of predictive validity studies of SRSS-IE scores in secondary schools).

As described in Lane et al. (2018), the district developed a secure system for teachers to complete the SRSS-IE independently. Students' names and district identification numbers were prepopulated for each elementary teacher's homeroom approximately 30 days, before each screening window opened. Principals were permissioned to view electronic folders prepared for each teacher two business days before teachers had electronic access to their individual folder. Several principals examined the folder structure for their schools to ensure each teacher had a folder and the correct students were prepopulated for each teachers' class. Teachers had electronic access to only their homeroom class. As part of their professional learning on systematic screening, teachers learned only two total scores, SRSS-E7 and SRSS-I5, would be used for decision making and not item-level data.

The Ci3T District Leadership Team provided de-identified student-level data electronically with principal investigators following Institutional Review Board and district-approved study procedures. In this article, we report results of fall 2015 elementary screening data in predicting spring 2016 year-end outcomes: ORF scores (Grades 1–5), MAP (Grades 1–5), nurse visits (Grades K–5), and in-school suspensions (Grades K–5).

Measures

SRSS-IE. The SRSS-IE is an efficient, free-access screening tool. Initial items included steal; lie, cheat, sneak; behavior problems; peer rejection; low academic achievement; negative attitude; and aggressive behavior. Results of a series of psychometric studies yielded five additional items to assess risk for internalizing behaviors: emotionally flat; shy, withdrawn; sad, depressed; anxious; lonely. Homeroom teachers completed the SRSS-IE for each student, rating each behavior on a 4-point, Likert-type scale developed by Drummond (1994) of *never* = 0, *occasionally* = 1, *sometimes* = 2, and *frequently* = 3. The original seven items were summed to form the SRSS-E7 score, with total scores used to place students into one of three risk groups: 0–3 low risk, 4–8 moderate risk, 9–21 high risk. The new five items were summed to form the SRSS-I5 score, with total scores used to place students into one of three risk groups: 0–1 low risk, 2–3 moderate risk, 4–15 high risk (Lane, Oakes, Swogger, et al., 2015). In this study, we used cut scores for SRSS-E7 and SRSS-I5 subscale scores established, respectively, by Drummond (1994) and Lane, Oakes, Swogger, et al. (2015) to examine predictive validity.

Extant schoolwide data. We predicted year-end outcomes: spring ORF scores, MAP, nurse visits, and in-school suspensions. Consistent with procedures described by Lane et al. (2018), district leaders provided de-identified year-end data electronically to principal investigators. ORF referred to the students' spring benchmark AIMSweb scores (number of words read correct per min). MAP referred to students' spring reading assessment percentile scores. Nurse visits referred to the number of visits a student made to the school nurse for assistance (e.g., getting a bandage, nausea, fever, somatic complaints). In-school suspensions referred to the number of days a student was assigned in-school suspension (a sanction reserved for serious rule infractions such as bullying). Each Ci3T Leadership Team developed a schoolwide reactive plan as part of their Ci3T model of prevention, listing, and defining behaviors warranting an in-school suspension. On receiving data, project staff conducted a series of logic checks to ensure data received reflected accurate ranges.

Statistical Analysis

Students were grouped into low-, moderate-, and high-risk levels as formerly described. Next, we examined potential group differences in ORF, MAP, nurse visits, and in-school suspensions. To explore potential differences in academic performance (ORF and MAP) by group, we fit a mixed-model analysis of variance (ANOVA) with group as a fixed effect and classroom teacher as a random effect to account for the nested nature of the data (students nested in teachers' classes; Lane et al., 2018). This model allowed us to examine the extent to which students scoring in the low-, moderate-, and high-risk categories according to fall SRSS-E7 and SRSS-I5 scores could be distinguished on spring ORF and MAP scores. Significant group effects were followed up with a set of pairwise comparisons ($k = 3$ comparisons: low vs. high, low vs. moderate, and moderate vs. high). We used a Bonferroni correction to adjust the Type I error rate for post hoc tests, with the alpha level for each group comparison set at 0.0167 (0.05/3).

Nurse visits and in-school suspensions were measured as counts. For these dependent variables, we computed a series of random-effects negative binomial regressions with an overdispersion parameter. These models account for the nested nature of the data (students nested in teachers' classes) when examining the degree to which students in low-, moderate-, and high-risk groups according to SRSS-E7 and SRSS-I5 scores could be differentiated on these year-end outcomes. As explained in Lane, Oakes, Swogger, et al. (2015), we fit negative binomial regression models for these outcome variables given their respective distributions closely resemble a Poisson distribution (commonly seen in count variables), with many students in the sample receiving zeros (e.g., zero nurse visits, zero in-school suspensions). The negative binomial regression model is useful when dependent variables are distributed as count data, and standard deviation exceeds the mean count (as with data presented here).

Analyses were computed using available data. While missing data were not imputed, missingness was managed using full maximum likelihood estimation for the mixed-model ANOVAs and negative binomial regressions (Enders, 2010).

We calculated effect sizes from observed means and standard deviations to determine the magnitude of differences between groups. We used Hedges' g formula, which incorporates the pooled standard deviation in the denominator and accounts for differences in the number of cases between groups. Effect sizes were interpreted per the following criteria: small- (0.20), medium- (0.50), and large-magnitude effects (0.80; Cohen, 1988).

Results

Externalizing: SRSS-E7

Findings of a mixed-model ANOVA with group as the between-participants fixed effect and teacher as the random effect indicated a group effect for ORF, $F(2, 620) = 39.51$, $p < .0001$ ($R^2 = 11\%$). The low-risk externalizing group earned statistically significantly higher ORF scores than moderate- (mean difference = 24.62, 95% confidence limits [16.05, 33.18], $t = 5.64$, $p < .0001$, Hedges' $g = 0.61$) and high-risk groups (mean difference = 47.41, 95% confidence limits [35.06, 59.76], $t = 7.54$, $p < .0001$, Hedges' $g = 1.18$). The moderate-risk group had a statistically significantly higher mean ORF than the high-risk group (mean difference = 22.80, 95% confidence limits [8.70, 36.89], $t = 3.18$, $p = .0016$, Hedges' $g = 0.52$, 95% confidence limits [8.70, 36.89]; see Table 3). Please see Table 4 for Pearson correlation coefficients.

Findings of a mixed-model ANOVA with group as the between-participants fixed effect and teacher as the random effect indicated a group effect for MAP reading percentile scores, $F(2, 2688) = 238.56$, $p < .0001$ ($R^2 = 15\%$). The low-risk externalizing group earned statistically significantly higher MAP scores than moderate- (mean difference = 23.63, 95% confidence limits [20.81, 26.45], $t = 16.45$, $p < .0001$, Hedges' $g = 0.87$) and high-risk groups (mean difference = 33.22, 95% confidence limits [29.23, 37.22], $t = 16.32$, $p < .0001$, Hedges' $g = 1.24$). The moderate-risk group had a statistically significantly higher mean MAP reading percentile score than the high-risk group (mean difference = 9.60, 95% confidence limits [5.01, 14.18], $t = 4.10$, $p < .0001$, Hedges' $g = 0.32$).

For number of nurse visits, we fit a random-effects negative binomial regression model. The model demonstrated a significant overall omnibus test, $F(2, 4244) = 123.20$,

Table 3. Elementary School: Behavioral and Academic Characteristics of Risk Groups According to Fall SRSS-IE Subscale Scores.

	Risk				Effect size		
	Low	Moderate	High				
Subscale/ variable	*M* (*SD*) *n*	*M* (*SD*) *n*	*M* (*SD*) *n*	Significance testing	L: M	L: H	M: H
Externalizing							
Oral reading fluency	163.23 (39.66) 468	138.62 (42.70) 107	115.82 (46.21) 46	L > M > H	0.61	1.18	0.52
MAP reading	66.54 (26.48) 2,047	42.91 (30.37) 443	33.32 (29.82) 199	L > M > H	0.87	1.24	0.32
Nurse visits	6.14 (6.81) 3,256	9.18 (9.59) 820	11.83 (9.89) 389	L < M < H	0.41	0.79	0.27
In-school suspensions	0.0052 (0.08) 3,256	0.0427 (0.30) 820	0.1080 (0.46) 389	L < M < H	0.23	0.63	0.20
Internalizing							
Oral reading fluency	159.04 (41.45) 459	150.59 (45.76) 88	139.18 (46.53) 74	L > H L = M; M = H	0.20	0.47	0.25
MAP reading	63.38 (28.32) 2,070	53.93 (32.15) 356	43.57 (30.47) 263	L > M > H	0.33	0.69	0.33
Nurse visits	6.84 (7.37) 3,387	7.59 (8.05) 628	9.33 (10.81) 450	L < M < H	0.10	0.32	0.19
In-school suspensions	0.0142 (0.15) 3,387	0.0510 (0.3580) 628	0.0311 (0.1978) 450	L < M, H M = H	0.20	0.13	0.07

Note. See text for 95% confidence limits. SRSS-IE = *Student Risk Screening Scale—Internalizing and Externalizing*; L = low risk; M = moderate risk; H = high risk; MAP = Measures of Academic Progress.

$p < .0001$. Post hoc comparisons revealed the low-risk-for-externalizing group experienced significantly fewer nurse visits than moderate- (mean difference = −0.40, 95% confidence limits [−0.47, −0.32], $t = -10.66$ $p < .0001$, Hedges' $g = 0.41$) and high-risk groups (mean difference = −0.68, 95% confidence limits [−0.78, −0.58], $t = -13.28$, $p < .0001$, Hedges' $g = 0.79$). The moderate-risk group had statistically higher mean scores for nurse visit compared with the high-risk group for externalizing behaviors (mean difference = −0.28, 95% confidence limits [−0.39, −0.17], $t = -4.91$, $p < .0001$, Hedges' $g = 0.27$).

For number of in-school suspensions, we fit a random-effects negative binomial regression model. The model demonstrated a significant overall omnibus test, $F(2, 4{,}244) = 41.44$, $p < .0001$. Post hoc comparisons revealed the low-risk-for-externalizing group experienced significantly fewer in-school suspensions than moderate- (mean difference = −2.18, 95% confidence limits [−2.84, −1.52], $t = -6.48$, $p < .0001$, Hedges' $g = 0.23$) and high-risk groups (mean difference = −3.27, 95% confidence limits [−3.99, −2.56], $t = -8.94$, $p < .0001$, Hedges' $g = 0.63$). Furthermore, students in the moderate-risk group earned fewer in-school suspensions than students in the high-risk groups (mean difference = −1.10, 95% confidence limits [−1.74, −0.46], $t = -3.35$ $p = .0008$, Hedges' $g = 0.20$).

Internalizing: SRSS-I5

Findings of a mixed-model ANOVA with group as the between-participants fixed effect and teacher as the random effect indicated a group effect for ORF, $F(2, 620) = 7.57$, $p = .0006$ ($R^2 = 2\%$). The low-risk internalizing group earned statistically significantly higher ORF scores than the high-risk group (mean difference = 19.87, 95% confidence limits [9.36, 30.37], $t = 3.71$, $p = .0002$, Hedges' $g = 0.47$). There were no statistically significant differences in ORF scores between low- and moderate- or between moderate- and high-risk groups (see Table 3).

Findings of a mixed-model ANOVA with group as the between-participants fixed effect and teacher as the random effect indicated a group effect for MAP reading percentile scores, $F(2, 2{,}688) = 63.74$, $p < .0001$ ($R^2 = 5\%$). The low-risk internalizing group earned statistically significantly higher MAP scores than moderate- (mean difference = 9.45, 95% confidence limits [6.18, 12.72], $t = 5.67$, $p < .0001$, Hedges' $g = 0.33$) and high-risk groups (mean difference = 19.81, 95% confidence limits [16.08, 23.54], $t = 10.41$, $p < .0001$, Hedges' $g = 0.69$). The moderate-risk group had a statistically significantly higher mean MAP reading percentile score than the high-risk group (mean difference =10.36, 95% confidence limits [5.72, 14.99], $t = 4.38$, $p < .0001$, Hedges' $g = 0.33$).

Table 4. Pearson Correlation Coefficients Between SRSS-IE and Outcome Measure Scores.

Variable	Externalizing *r* *p* value *N*	Internalizing *r* *p* value *N*	Oral reading fluency *r* *p* value *N*	MAP reading *r* *p* value *N*	Nurse visits *r* *p* value *N*	In-school suspensions *r* *p* value *N*
Externalizing	1.00 4,594	0.334594 <.0001 4,594	−0.33662 <.0001 621	−0.38097 <.0001 2,689	0.23284 <.0001 4,465	0.15147 <.0001 4,465
Internalizing	0.33025 <.0001 4,594	1.00 4,594	−0.15420 <.0001 621	−0.21281 <.0001 2,689	0.09400 <.0001 4,465	0.04806 .0013 4,465
Oral reading fluency	−0.33662 <.0001 621	−0.15420 <.00,001 621	1.00 621	0.78409 <.0001 601	−0.19581 <.0001 604	−0.19581 .2057 604
MAP reading	−0.38097 <.0001 2,689	−0.21281 <.0001 2,689	0.78409 <.0001 601	1.00 2,689	−0.19269 <.0001 2,689	−0.06383 .0009 2,689
Nurse visits	0.23284 <.0001 4,465	0.09400 <.0001 4,465	−0.19581 <.0001 604	−0.19269 <.0001 2,689	1.00 4,465	0.10099 <.0001 4,465
In-school suspensions	0.15147 <.0001 4,465	0.04806 .0013 4,465	−0.05156 .2057 604	−0.06383 .0009 2,689	0.10099 <.0001 4,465	1.00 4,465

Note. SRSS-IE = *Student Risk Screening Scale—Internalizing and Externalizing; r* = Pearson correlation coefficient; MAP = Measures of Academic Progress.

For number of nurse visits, we fit a random-effects negative binomial regression model. The model demonstrated a significant overall omnibus test, $F(2, 4{,}244) = 17.14$, $p < .0001$. Post hoc comparisons revealed the low-risk-for-internalizing group experienced significantly fewer nurse visits than moderate- (mean difference = −0.11, 95% confidence limits [−0.19, −0.02], $t = -2.54$ $p = .0111$, Hedges' $g = 0.10$) and high-risk groups (mean difference = −0.29, 95% confidence limits [−0.39, −0.19], $t = -5.64$, $p < .0001$, Hedges' $g = 0.32$). The moderate-risk group had statistically higher mean scores for nurse visit compared with the high-risk group for internalizing behaviors (mean difference = −0.18, 95% confidence limits [−0.30, −0.06], $t = -2.94$, $p = .0033$, Hedges' $g = 0.19$).

For number of in-school suspensions, we fit a random-effects negative binomial regression model. The model demonstrated a significant overall omnibus test, $F(2, 4{,}244) = 9.35$, $p < .0001$. Post hoc comparisons revealed the low-risk-for-internalizing group experienced significantly fewer in-school suspensions than moderate- (mean difference = −1.36, 95% confidence limits [−2.02, −0.70], $t = -4.05$, $p < .0001$, Hedges' $g = 0.20$) and high-risk groups (mean difference = −1.08, 95% confidence limits [−1.91, −0.25], $t = -2.55$, $p = .0109$, Hedges' $g = 0.13$). There were no statistically significant differences in mean in-school suspension scores between moderate- and high-risk groups.

Discussion

Psychometric studies of SRSS-IE scores offer evidence of reliability and validity, including results suggesting comparable accuracy between SRSS-IE scores and other validated screening tools' scores (e.g., SSBD, SDQ, SSiS-PSG) in detecting students with externalizing and internalizing behaviors. We conducted the present study to explore predictive validity of SRSS-IE scores, an essential next step in the systematic line of inquiry establishing the SRSS-IE as a psychometrically sound, feasible tool for school use.

Findings offer additional evidence of fall SRSS-E7 scores predicting behavioral and academic year-end outcomes for elementary-age students consistent with early inquiry of the original SRSS tool (e.g., Menzies & Lane, 2012; Oakes et al., 2010). Furthermore, we provide initial evidence suggesting fall SRSS-I5 scores also predict important educational outcomes for students.

Predictive Validity of Externalizing Scores in Elementary Schools

In predicting year 1 outcomes at the elementary school level, fall SRSS-E7 scores differentiated low-, moderate-, and high-risk groups on ORF, MAP reading, nurse visits, and in-school suspensions. The low-risk group had

statistically significantly higher year-end ORF and MAP scores, and fewer nurse visits and in-school suspensions compared with moderate- and high-risk groups. Furthermore, moderate- and high-risk group mean scores could also be differentiated, with students in the high-risk group having the most negative outcomes.

Findings were highly similar to previous short-term (1 year) predictive validity studies indicating fall SRSS-E7 scores predicted year-end behavioral outcomes (e.g., ODRs and self-control skills; Menzies & Lane, 2012; Oakes et al., 2010), and academic outcomes, such as ORF (Oakes et al., 2010) and proficiency in language art skills (Menzies & Lane, 2012). Yet this is the first study at the elementary level to explore the extent to which screening scores predicted in-school suspensions and nurse visits. We learned fall SRSS-E7 scores differentiated students in the low-, moderate-, and high-risk for externalizing behaviors with students in the low-risk group having fewer nurse visits and fewer days spent in in-school suspension than students in the moderate- and high-risk groups. Also, students in the moderate-risk group had fewer nurse visits and fewer days spent in in-school suspension than students in the high-risk group. Given this is the first study examining these variable, the information should be considered preliminary until these findings are replicated to be certain results are not spurious (Cook, 2014). However, a recent psychometric study of the SRSS-IE in middle and high schools suggested fall SRSS-E7 scores could differentiate secondary school nurse visits (Lane et al., 2018). As discussed in that article, future inquiry into nurse visits is warranted as frequent visits to the nurse could signal a range of concerns. For example, it may be students use visits to the nurse's office to manage issues such as anxiety (e.g., panic attacks), the need to escape too difficult or too easy instruction, somatic complaints, or a desire to seek solace or attention from an adult in a helping profession. Students involved in physical aggression and altercations might make repeated visits as well. Conversely, other students may visit the nurse for medication management. In short, frequent trips to the nurse can indicate a range of needs, which will vary from student to student (Johnson & Hutcherson, 2006; Vernberg, Nelson, Fonagy, & Twemlow, 2011). The same is true for in-school suspensions (although the base rate was very low). Behaviors leading to in-school suspension may indicate an unmet need, and offering students additional, and proactive, supports (often in the form of evidence-based Tier 2 and 3 supports) may reduce nurse visits or suspensions. Limiting nurse visits to those with a medical need would enable nurses to better manage the health needs of students and reduce overall burden. For example, the National Association of School Nurses recommends a nurse-to-student ratio of 1:750 in a healthy context and a lower ratio in contexts in which students have more nuanced health needs. This recommended ratio is frequently not obtained due to shortages of school nurses and/or fiscal challenges (American Association of Colleges of Nursing, 2014; AAP Council on School Health, 2016).

In examining effect sizes, results indicated medium-to-large magnitude effects when differentiating low- and high-risk groups on ORF (1.18), MAP (1.24), nurse visits (0.79), and in-school suspensions (0.63). Effect sizes were medium-to-large when differentiating ORF and MAP scores between low- and moderate-risk groups (0.61 and 0.87, respectively), yet small-to-moderate for nurse visits (0.41) and in-school suspensions (0.23). Collectively, results suggest fall SRSS-E7 scores continue to be an effective screening tool for predicting behavioral and academic outcomes for elementary-age students.

Predictive Validity of Internalizing Scores in Elementary Schools

In this first predictive validity study of SRSS-I5 scores in elementary schools, results suggested kindergarten through fifth-grade students at low, moderate, and high risk for internalizing behaviors could also be differentiated on MAP reading and nurse visits for all three risk groups as was the case with fall SRSS-E7 scores. Students in the low-risk category had higher mean MAP scores and fewer mean nurse visits than students in the moderate- and high-risk groups. Students in the moderate-risk group had higher mean MAP scores and fewer mean nurse visits than students in the high-risk groups. The distinction between the three groups was very clear on these two variables as was the case with fall SRSS-E7 scores, yet the magnitude of these differences was smaller.

For ORF and in-school suspensions, the low-risk group could be differentiated from the high-risk group in each case (with the low-risk group experiencing more favorable outcomes). While the low-risk group also had fewer mean in-school suspensions, ORF scores did not distinguish between low- and moderate- or moderate- and high-risk groups. Students in the moderate-risk group had mean scores similar to those in the low- and high-risk group; thus, students with internalizing concerns may not demonstrate detectable differences in reading progress until the internalizing concerns reach the criteria for high risk. These findings highlight the complexity of the internalizing behaviors and school outcomes for students with and at risk for internalizing behaviors. Internalizing behaviors reflect a broad array of more covert behavioral manifestations such as anxiety, social withdrawal, and depression (Bradshaw et al., 2008; Green et al., 2017). As discussed for several decades, students with strong interpersonal skills are able to interact comfortably with a range of individuals: peers, teachers, parents, and other authority figures (Rapport, Denney, Chung, & Hustace, 2001; Walker et al., 2014). Yet students who experience internalizing behaviors often struggle in

these important relationships, making school engagement challenging, and academic outcomes may be affected for those with the highest levels of teacher-rated risk.

Given the lack of clear distinction between moderate- and high-risk groups on several variables, it will be important to ensure *any* indication of risk for internalizing behaviors be attended to swiftly. Educators may consider the use of low-intensity teacher supports (e.g., intentional use of behavior-specific praise to acknowledge and engage students) or more targeted supports according to students' needs (Lane et al., 2018).

When considering effect sizes, it should be noted the magnitude of differences between low- and high-risk internalizing groups were lower than the differences between externalizing groups for all variables. These small-magnitude differences between groups on in-school suspensions were expected. This finding was comparable with results reported by McIntosh, Campbell, Carter, and Zumbo (2009) whose inquiry also suggested students with internalizing concerns are not adequately detected through reactive procedures such as ODRs that may result in in-school suspensions at the elementary level.

As discussed, replication is essential before drawing a definitive conclusion regarding the predictive validity of SRSS-I5 scores in predicting academic and behavioral outcomes. In the interim, we urge caution as replication is needed before these results are generalized to other locales and with students from other contexts (e.g., more diverse backgrounds). At the same time, it would not be wise to prematurely conclude risk is simply a dichotomous variable (low vs. any risk) when examining outcomes. Fall SRSS-I5 scores may suggest any level of risk at the onset of an academic year is cause for concern and may warrant additional consideration or support depending on the breadth of concern (Lane, Oakes, Ennis, & Royer, 2015; Walker et al., 2014).

Educational Implications

We are pleased to offer findings from this researcher–practitioner partnership as additional information on the utility of SRSS-E7 scores and the preliminary nature of the utility of SRSS-I5 scores. Results suggested SRSS-IE scores are useful for distinguishing elementary students in the low-risk group from students in the moderate- and high-risk groups on most academic and behavioral variables examined in this study. Although often difficult to detect, the fall SRSS-I5 scores did distinguish—at a minimum—between students in low- and high-risk groups on all variables.

With the recommendation that systematic screenings take place three times each year (fall, winter, and spring), it will be important to explore the degree to which additional time with students (e.g., winter and/or spring scores) will increase predictive validity of SRSS-E7 and SRSS-I5 scores collected at later time points. Studies of SRSS-E7 scores suggest this is indeed the case at the high school level (Lane, Oakes, Ennis et al., 2013), and we hypothesize this will be the case at the elementary level. A key consideration is the need to determine whether winter and spring internalizing scores are more accurate in predicting student outcomes 1 year later. Given the host of negative outcomes of these difficult-to-detect and often more covert behaviors, this is a key point for future inquiry.

At this time, we encourage school leadership teams and individual teachers to move forward cautiously when utilizing fall screening scores. In optimal conditions, screening scores should be examined in conjunction with other reliable, available data to shape instruction. When working in tiered systems, multiple sources of data can be used to inform Tier 1 practices, teacher-delivered practices, and the use of Tiers 2 and 3 for students with targeted and intensive intervention needs, respectively. For example, in schools in which more than 20% of students are placing into moderate- and high-risk categories for externalizing or internalizing behaviors, instructional coaches might offer professional learning to all classified and certified staff and parents in validated strategies such as behavior-specific praise and incorporating choice (instructional choice in schools and choice of activities in the home settings; Royer, Lane, Dunlap, & Ennis, 2018) as a Tier 1 practice given the magnitude of the students experiencing behavioral risk. Then, at the next screening time point, students still scoring in the moderate-risk category despite high-fidelity implementation of Tier 1 practices might be supported with self-management strategies (Carter, Lane, Crnobori, Bruhn, & Oakes, 2011) or cognitive restructuring activities as Tier 2 practices (Smith, Taylor, Barnes, & Daunic, 2012). Support at each level of prevention will require high-quality professional learning, ideally with positive practice, coaching, and performance feedback.

Just as explained in the recent IES RFA, the intent is to examine the potential benefits of tiered systems as a mechanism that "provides academic, social, emotional, and behavioral support for all students, and provides resources and supports that teachers and other school personnel need to support" students with and at risk for learning and behavioral challenges in authentic educational settings (US Department of Education, IES, 2017, p. 17). As mentioned, systematic screening efforts will play a central role in focusing intervention efforts. In short, one goal will be to empower faculty, staff, and parents with the knowledge, skills, and confidence to incorporate positive behavior interventions and supports as part of daily activities. It is important to move past the erroneous idea that academic, behavior, and social competencies should be addressed in isolation rather than simultaneously (McIntosh & Goodman, 2016; Menzies, Lane, Oakes, & Ennis, 2017). Results from this study suggest soft-signs of externalizing

Table 5. Co-Occurrence Between Internalizing and Externalizing Behaviors Measured by the SRSS-IE.

Externalizing risk category (SRSS-E)	Internalizing category (SRSS-I)			
	Low	Moderate	High	Total
Low	2,773[a]	383	192	3,348
	60.36[b]	8.34	4.18	72.88
	82.83[c]	11.44	5.73	
	79.62[d]	59.20	41.38	
Moderate	556	160	131	847
	12.10	3.48	2.85	18.44
	65.64	18.89	15.47	
	15.96	24.73	28.23	
High	154	104	141	399
	3.35	2.26	3.07	8.69
	38.60	26.07	35.34	
	4.42	16.07	30.39	
Total	3,483	647	464	4,594
	75.82	14.08	10.10	100.00

Note. SRSS = *Student Risk Screening Scale*; E = externalizing; I = internalizing.
[a]Frequency. [b]Percentage. [c]Row percentage. [d]Column percentage.

and internalizing behaviors in the elementary years predict important academic and behavioral outcomes. We hope this information is useful as we move forward with a comprehensive, integrated approach to meeting students' multiple needs.

Limitations and Future Directions

We encourage readers to interpret results with attention to the following limitations. First, as with all studies, replication is essential before generalizing findings (Cook, 2014; Travers, Cook, Therrien, & Coyne, 2016). Specifically, although this study included a large sample of students, they were from one district in one geographical region. As recommended by NCII Behavior Screening TRC, it is necessary for additional inquiry in other geographical locales and ideally with ethnically and culturally diverse samples. This is particularly important when interpreting SRSS-I5 scores as this is the first study examining predictive validity of SRSS-I5 scores in predicting elementary students' academic and behavioral outcomes.

Second, in this study, we analyzed SRSS-E7 and SRSS-I5 scores in isolation. We encourage other research teams to explore issues of comorbidity by examining predictive validity of combined subscale score (e.g., total scores) given the fact students often present with externalizing and internalizing behavior patterns, as was the case with the present sample (see Table 5). For example, it would be interesting to examine the extent to which students with various facets of EBD (e.g., externalizing but not internalizing, internalizing but not externalizing, and co-occurring challenges) fared over time in schools implementing a tiered system of supports. Although not the focus of this study, it would be interesting to note whether students accessed evidence-based strategies, practices, and programs as Tier 2 and 3 supports and how they responded to this extra assistance. This study examined the constructs of externalizing and internalizing in isolation without attention to issues of comorbidity. While this was not a goal of this predictive validity study, we encourage other research teams to examine the predictive validity of SRSS-IE scores when used in tandem to address comorbidity. For example, an important next step in this line of inquiry is to determine the degree to which co-occurrence of externalizing and internalizing patterns (e.g., students with intensive intervention needs for externalizing and internalizing behaviors) predicts important educational outcomes for students (e.g., academic performance and behavioral and social performance patterns; Lane et al., 2018).

Third, we encourage replication with larger samples to examine academic outcomes. Also, ORF data were available on more than 600 students; this was a very small percentage of the present sample. Schools in this sample were in their first year of collecting these data, and the practice had not yet been taken to scale districtwide. We emphasize all available data collected by the district were analyzed.

Fourth, the elementary schools in the present study were supported as part of researcher–practitioner partnership. The Ci3T District Leadership Team adopted screening as part of their districtwide implementation of Ci3T. They built a screening platform managed by the instructional technology departments in conjunction with their teaching and learning department. Thus, the district-certificated employees received ongoing professional learning on how to conduct and utilize data gleaned from systematic screening to inform instruction. While these features are a clear strength, future inquiry is needed to determine whether these findings are replicated in other systems where systematic screening is supported without the additional resources of research–partnerships (Lane et al., 2018; Lane, Oakes, Ennis et al., 2013).

Summary

Despite limitations, results of this psychometric study of SRSS-IE scores offer evidence of predictive validity of SRSS-E7 (measuring externalizing behaviors) and SRSS-I5 (measuring internalizing behaviors) scores for predicting a range of academic and behavioral outcomes for elementary students. Results suggest students with high levels of risk (particularly those with externalizing behaviors) as measured by the SRSS-IE at the fall administration time point were more apt to have lower ORF scores, and have more nurse visits compared with students at low risk according to SRSS-IE scores. We encourage readers to avoid generalization errors and use this information cautiously until

replication studies confirm these findings (Cook, 2014; Travers et al., 2016). Yet this study presents important findings in this programmatic line of inquiry offering evidence that one teacher's independent rating in the fall can differentiate between elementary students with low and high risk for both major disorders of childhood (externalizing and internalizing behaviors) on behavioral (proximal) and academic (distal) measures.

Authors' Note

Opinions expressed herein are those of the authors and do not necessarily reflect the position of the U.S. Department of Education, and such endorsements should not be inferred.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Office of Special Education Programs U.S. Department of Education (H326S980003) and Institute of Education Sciences Partnership Grant (R305H150018).

ORCID iDs

Kathleen Lynne Lane https://orcid.org/0000-0001-6364-838X
David J. Royer https://orcid.org/0000-0003-2882-1049
Mark Matthew Buckman https://orcid.org/0000-0001-9332-0940
Grant Edmund Allen https://orcid.org/0000-0002-9743-3368

References

AAP Council on School Health. (2016). Role of the school nurse in providing school health services. *Pediatrics, 137*(6), e20160852. doi:10.1542/peds.2016-0852

American Association of Colleges of Nursing. (2014). *Nursing shortage*. Retrieved from http://www.aacn.nche.edu/media-relations/fact-sheets/nursing-shortage

Bradshaw, C. P., Buckley, J. A., & Ialongo, N. S. (2008). School-based service utilization among urban children with early onset educational and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly*, *23*, 169–186. doi:10.1037/1045-3830.23.2.169

Carter, E. W., Lane, K. L., Crnobori, M. E., Bruhn, A. L., & Oakes, W. P. (2011). Self-determination interventions for students with and at risk for emotional and behavioral disorders: Mapping the knowledge base. *Behavioral Disorders*, *36*, 100–116.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cook, B. G. (2014). A call for examining replication and bias in special education research. *Remedial and Special Education*, *35*, 233–246. doi:10.1177/0741932514528995

Cook, B. G., & Tankersley, M. (Eds.). (2013). *Effective practices in special education*. Boston, MA: Pearson.

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

Elliott, S. N., & Gresham, F. (2008). *Social Skills Improvement System (SSiS)—Performance screening guide*. San Antonio, TX: PsychCorp Pearson Education.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

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

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 1337–1345. doi:10.1097/00004583-200111000-00015

Green, J. G., Comer, J. S., Donaldson, A. R., Elkins, R. M., Nadeau, M. S., Reid, G., & Pincus, D. B. (2017). School functioning and use of school-based accommodations by treatment-seeking anxious children. *Journal of Emotional and Behavioral Disorders*, *25*, 220-232. doi:10.1177/1063426616664328

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. 1400 et esq. (2004).

Johnson, V., & Hutcherson, V. (2006). A study of the utilization patterns of an elementary school-based health clinic over a 5-year period. *Journal of School Health*, *76*, 373–378. doi:10.1111/j.1746-1561.2006.00128.x

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior Assessment System for Children-Third edition (BASC-3): Behavioral and Emotional Screening System (BESS)*. Bloomington, MN: Pearson.

Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & von der Embse, N. P. (2013). *Social, Academic, and Emotional Behavior Risk Screener©: Teacher rating form*. Columbia: University of Missouri.

Lane, K. L., Carter, E., Jenkins, A., Magill, L., & Germer, K. (2015). Supporting comprehensive, integrated, three-tiered models of prevention in schools: Administrators perspectives. *Journal of Positive Behavior Interventions*, *17*, 209–222. doi:10.1177/1098300715578916

Lane, K. L., Kalberg, J. R., Lambert, W., Crnobori, M., & Bruhn, A. (2010). A comparison of systematic screening tools for emotional and behavioral disorders: A replication. *Journal of Emotional and Behavioral Disorders*, *18*, 100–112. doi:10.1177/1063426609341069

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., Little, M. A., Casey, A. M., Lambert, W., Wehby, J. H., Weisenbach, J. L., & Phillips, A. (2009). A comparison of systematic screening tools for emotional and behavioral disorders: How do they compare? *Journal of Emotional and Behavioral Disorders*, *17*, 93–105. doi:10.1177/1063426608326203

Lane, K. L., & Menzies, H. M. (2009). *Student Risk Screening Scale for Internalizing and Externalizing (SRSS-IE)*. Retrieved from http://www.ci3t.org/screening

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., Cantwell, E. D., & Royer, D. J. (2016). *Building and installing comprehensive, integrated, three-tiered (Ci3T) models of prevention: A practical guide to supporting school success*. Phoenix, AZ: KOI Education.

Lane, K. L., Oakes, W. P., Cantwell, E. D., Royer, D. J., Leko, M., Schatschneider, C., & Menzies, H. M. (2018). Predictive validity of Student Risk Screening Scale for Internalizing and Externalizing (SRSS-IE) scores in secondary schools. *Journal of Emotional and Behavioral Disorders*. Advance online publication. doi: 10.1177/1063426617744746

Lane, K. L., Oakes, W. P., Common, E. A., Zorigian, K., Brunsting, N. C., & Schatschneider, C. (2015). A comparison between SRSS-IE and SSiS-PSG scores: Examining convergent validity. *Assessment for Effective Intervention*, *40*, 114–126. doi:10.1177/1534508414560346

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. doi:10.1177/1063426611407339

Lane, K. L., Oakes, W. P., Ennis, R. P., & Royer, D. J. (2015). Additional evidence of convergent validity between SRSS-IE and SSiS-PSG scores. *Behavioral Disorders*, *40*, 213–229. doi:10.17988/0198-7429-40.4.213

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., Oakes, W. P., Swogger, E. D., Schatschneider, C., Menzies, H. M., & Sanchez, J. (2015). Student Risk Screening Scale for internalizing and externalizing behaviors: Preliminary cut scores to support data-informed decision making. *Behavioral Disorders*, *40*, 159–170. doi:10.17988/0198-7429-40.3.159

Lane, K. L., Richards-Tutor, C., Oakes, W. P., & Connor, K. (2014). Initial evidence for the reliability and validity of the Student Risk Screening Scale with elementary age English learners. *Assessment and Effective Intervention*, *39*, 219–232. doi:10.1177/1534508413496836

Lane, K. L., & Walker, H. M. (2015). The connection between assessment and intervention: How can screening lead to better interventions? In B. Bateman, M. Tankersley, & J. Lloyd (Eds.), *Enduring issues in special education: Personal perspectives* (pp. 283–301). New York, NY: Routledge.

Maggin, D. M., Wehby, J. H., Farmer, T. W., & Brooks, D. S. (2016). Intensive interventions for students with emotional and behavioral disorders: Issues, theory, and future directions. *Journal of Emotional and Behavioral Disorders*, *24*, 127–137. doi:10.1177/1063426616661498

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior supports. *Behavioral Disorders*, *34*, 100–113.

McIntosh, K., & Goodman, S. (2016). *Integrating multi-tiered systems of support: Blending RTI and PBIS*. New York, NY: Guilford Press.

Menzies, H. M., & Lane, K. L. (2012). Validity of the Student Risk Screening Scale: Evidence of predictive validity in a diverse, suburban elementary setting. *Journal of Emotional and Behavioral Disorders*, *20*, 82–91. doi:10.1177/1063426610389613

Menzies, H. M., Lane, K. L., Oakes, W. P., & Ennis, R. P. (2017). Increasing students' opportunities to respond: A strategy for supporting engagement. *Intervention in School and Clinic*, *52*, 204–209. doi:10.1177/1053451216659467

National Center on Intensive Intervention. (2017). *Behavior screening rating rubric*. American Institutes for Research. Retrieved from http://www.intensiveintervention.org/sites/default/files/NCII_BScreening_RatingRubric_July2017.pdf

New Teacher Project. (2013). *Perspectives of irreplaceable teachers: What America's best teachers think about teaching*. Retrieved from http://tntp.org/publications/view/perspectives-of-irreplaceable-teachers-best-teachers-think-about-teaching

Oakes, W. P., Lane, K. L., Cox, M., & Messenger, M. (2014). Logistics of behavior screenings: How and why do we conduct behavior screenings at our school? *Preventing School Failure*, *58*, 159–170. doi:10.1080/1045988X.2014.895572

Oakes, W. P., Wilder, K., Lane, K. L., Powers, L., Yokoyama, L., O'Hare, M. E., & Jenkins, A. B. (2010). Psychometric properties of the Student Risk Screening Scale: An effective tool for use in diverse urban elementary schools. *Assessment for Effective Intervention*, *35*, 231–239. doi:10.1177/1534508410379796

Rapport, M. D., Denney, C. B., Chung, K. M., & Hustace, K. (2001). Internalizing behavior problems and scholastic achievement in children: Cognitive and behavioral pathways as mediators of outcome. *Journal of Clinical Child & Adolescent Psychology*, *30*, 536–551. doi:10.1207/s15374424jccp3004_10

Royer, D. J., Lane, K. L., Dunlap, K., & Ennis, R. P. (2018). A systematic review of teacher-delivered behavior-specific praise on K-12 student performance. *Remedial and Special Education*. Advance online publication. doi: 0741932517751054

Smith, S. W., Taylor, G. G., Barnes, T., & Daunic, A. P. (2012). Cognitive-behavioral interventions to prevent aggression of students with emotional and behavioral disorders. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *Advances in learning and behavioral disabilities: Vol. 25. Classroom behavior, contexts, and interventions* (pp. 47–70). Bingley, UK: Emerald Publishing.

Travers, J. C., Cook, B. G., Therrien, W. J., & Coyne, M. D. (2016). Replication research and special education. *Remedial and Special Education*, *37*, 195–204. doi:10.1177/0741932516648462

US Department of Education, Institute of Education Sciences. (2017). *Request for applications: Research networks focused on critical problems of policy and practice in special education: Multi-tiered systems of support* (CFDA No. 84.324N).

Vernberg, E. M., Nelson, T. D., Fonagy, P., & Twemlow, S. W. (2011). Victimization, aggression, and visits to the school nurse for somatic complaints, illnesses, and physical injuries. *Pediatrics*, *127*, 842–848. doi:10.1542/peds.2009-3415

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

Yudin, M. (2014). *PBIS: Providing opportunity. A keynote address presented at the National PBIS Leadership Forum*. Rosemont, IL: Reform.