

## Training Teachers to Facilitate Early Identification of Mental and Behavioral Health Risks

Nathaniel P. von der Embse, Stephen P. Kilgus, Katie Eklund, Ethan Ake & Shana Levi-Neilsen |

To cite this article: Nathaniel P. von der Embse, Stephen P. Kilgus, Katie Eklund, Ethan Ake & Shana Levi-Neilsen | (2018) Training Teachers to Facilitate Early Identification of Mental and Behavioral Health Risks, School Psychology Review, 47:4, 372-384, DOI: [10.17105/SPR-2017-0094.V47-4](https://doi.org/10.17105/SPR-2017-0094.V47-4)

To link to this article: <https://doi.org/10.17105/SPR-2017-0094.V47-4>



Published online: 27 Dec 2019.



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



Article views: 96



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

# Training Teachers to Facilitate Early Identification of Mental and Behavioral Health Risks

Nathaniel P. von der Embse  
*University of South Florida*

Stephen P. Kilgus  
Katie Eklund  
*University of Wisconsin-Madison*

Ethan Ake  
Shana Levi-Neilsen  
*Temple University*

*Abstract.* Universal screening has been proposed as a method for effectively and efficiently identifying students at risk for mental health concerns. As with traditional rating scales, screeners have historically been used in the absence of any user training. Recent research suggests, however, that systematic training can be used to enhance the psychometric defensibility of teacher rating data. Thus, the present investigation evaluated the effectiveness of a pilot teacher training intervention on improving identification of students with behavioral or emotional concerns. A total of 91 teachers (57 training, 34 control) participated in the study. Participants in the training group received information on the utility of universal screening and were provided with information regarding mental and behavioral risk in schools. Videos depicting examples of problem behaviors and student concerns were also provided, and trainees were given the opportunity to practice rating videos with performance feedback. Following training, teachers completed the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) for their 1,158 students in four urban elementary schools. Teachers also completed measures of perceived assessment usability and willingness and efficacy to engage in universal screening. Finally, schools provided extant data regarding each student's academic and behavioral functioning (e.g., benchmark assessments, suspensions, and disciplinary referrals). Structural equation modeling and moderation analyses indicated a stronger relationship between SAEBRS fall risk and end-of-year behavioral outcomes in the trained group, and this relationship was moderated by assessment acceptability. Implications for practice and future research are discussed.

*Keywords:* teacher training, universal screening, mental and behavioral health risk, assessment

Recent estimates have suggested that almost 25% of children will have a mental or behavioral health problem in a given year (Patel, Flisher, Hetrick, & McGorry, 2007; Perou et al., 2013). These problems can increase in severity if left untreated and lead to multiple detrimental impacts on child outcomes (Suldo, Gormley, DuPaul, & Anderson-Butcher, 2014). Many children do not have access to community-based mental health providers to address their needs and instead rely on school-based services (Bruhn, Woods-Groves, & Huddle, 2014; Roeser, van der Wolf, & Strobel, 2001). Urban schools,

in particular, may need significant support for mental and behavioral health services, with estimates on the percentage of students at risk ranging from 50% (Adelman & Taylor, 2010) to nearly 75% of students in another study (von der Embse, Rutherford, Mankin, & Jenkins, 2018). Thus, schools may be the ideal setting for early identification that would facilitate prevention and intervention efforts.

School-based mental health programs (SBMH) have been supported in federal initiatives, including the New Freedom Commission on Mental Health (2003) and

---

*Authors' Note.* The first and second authors have a financial interest in one of the measures discussed in this manuscript.

Correspondence regarding this article should be addressed to Nathaniel von der Embse, School Psychology, University of South Florida, Tampa, FL 33620; e-mail: natev@usf.edu

Copyright 2018 by the National Association of School Psychologists. ISSN 0279-6015, eISSN 2372-966x

have included a number of new evidenced-based programs (Franklin, Kim, Ryan, Kelly, & Montgomery, 2012). The efficacy of SBMH programs has led to widespread implementation (Clayton, Ballif-Spanvill, & Hunsaker, 2001) and a focus on scaling up across settings (Domitrovich & Greenberg, 2000). Concurrently, reauthorizations of educational laws have emphasized that schools should employ a public health framework (e.g., response to intervention [RTI], positive behavioral interventions and supports [PBIS]) that promotes early identification and intervention using a three-tiered model (Hoover, Baca, Wexler-Love, & Saenz, 2008). These frameworks, now integrated as multitiered systems of support (MTSS), have increased the involvement of general education teachers in the assessment and treatment of at-risk students and expedited the development and implementation of universal (or Tier 1) strategies (Franklin et al., 2012). However, there are substantial limitations on the ability of existing systems to serve the wide range of emotional and behavioral health needs of children. For instance, research suggests schools are underidentifying students with internalizing problems, such as those related to risk for depression or anxiety (McIntosh, Ty, & Miller, 2014). The primary goal of the present investigation was to examine the efficacy of a pilot teacher training program in facilitating early identification of students with emotional and behavioral problems.

### Identifying Risks for Emotional and Behavioral Problems in Schools

Traditional methods for identifying emotional and behavioral health risk in schools (e.g., teacher referral) have significant limitations and typically identify only a subset of students who need services (Kamphaus, DiStefano, Dowdy, Eklund, & Dunn, 2010). Many of these identification methods are reactive in nature and pose a number of challenges to the effective implementation of MTSS. For instance, office discipline referrals (ODRs) only identify approximately 20% of students at risk when compared to other, more systematic, methods (Miller et al., 2015). Furthermore, methods such as teacher referral and ODRs are subject to various biases (e.g., perceptions of teachability of certain types of students; Gerber & Semmel, 1984), often resulting in overidentification of certain student subgroups or behaviors (e.g., minority children with externalizing behavior problems; Skiba, Michael, Nardo, & Peterson, 2002) and underidentification of others (e.g., internalizing problems; Kahlberg, Lane, Driscoll, & Wehby, 2011).

Waiting for a sufficient accumulation of severe mental and behavioral health problems to warrant a comprehensive, individualized evaluation results in missed opportunities for early intervention (Oakes, Lane, Cox, & Messenger, 2014). Delays from initial problem manifestation to service delivery have ranged from 1 to 3 years for academics and nearly 5 to 7 years in the behavioral domain (Lloyd, Kauffman, Landrum, & Roe, 1991; Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007). Universal screening represents a proactive

alternative that may enable the identification of early symptoms indicative of later severe mental and behavioral health risk (Kamphaus, 2012). Teachers play a crucial role in identifying risk as part of universal screening procedures to facilitate subsequent early intervention supports.

### Role of Teachers in Facilitating Emotional and Behavioral Identification

There is a growing literature base supporting teacher influences on student emotional and behavioral outcomes (Lynn, McKay, & Atkins, 2003). Given the total time spent in the classroom with students, teachers are in a vital position to detect and provide support for students experiencing mental health problems (Johnson, Eva, Johnson, & Walker, 2011). For example, most elementary teachers spend over 1,000 h a year (6 h a day  $\times$  180 days a year) with a classroom of approximately 30 students. Equipping teachers with greater knowledge of and tools to detect risk of mental health problems represents an opportunity for improved service delivery. In a recent review, most SBMH programs were delivered at Tier 1 and included teachers approximately 40% of the time (Franklin et al., 2012).

However, research suggests teachers struggle to identify and refer students at risk of behavioral and emotional concerns (Severson et al., 2007). For example, teachers can be inconsistent with referral practices, with a recent study indicating that only 38% of students with emotional and behavioral health needs were referred for intervention (Eklund & Dowdy, 2014). Additionally, students with externalizing concerns are referred at a much higher rate than those with internalizing behaviors, suggesting only certain types of risk might be recognized at the expense of others (Pearcy, Clopton, & Pope, 1993). Universal screening, or the use of brief assessment instruments to evaluate a population for the presence of a characteristic of interest (e.g., behavioral risk; Jenkins, Hudson, & Johnson, 2007), has been increasingly adopted by schools to facilitate teacher-initiated early identification of a broad range of student risks. School-wide screening practices have become more prevalent within MTSS frameworks given the natural connection to tiered intervention and are a recommended practice by the National Association of School Psychologists (NASP), the National Research Council, and the Institute of Medicine (O'Connell, Boat, & Warner, 2009). Universal screening tools have been found to help teachers reliably identify students at risk in a more expedient manner while also informing the type and intensity of intervention (Vannest, 2012).

### Barriers to Screening Implementation

Although teachers have reported that universal screening provides useful information, there are barriers that may preclude the implementation or benefit of screening procedures (Greer, Wilson, DiStefano, & Liu, 2012). Although several barriers have been noted (Dever, Raines, & Barclay,

2012), we review two particularly salient concerns. First, research suggests that although screening tools are strong predictors of other rating scales, they are relatively inferior predictors of ecologically valid outcomes, defined as variables commonly emphasized by schools and considered as part of applied decision making and program evaluation. Within schools, such outcomes typically include both academic (e.g., curriculum-based measures [CBM], statewide achievement tests) and behavioral variables (e.g., ODRs, suspensions, attendance). A recent meta-analysis by Kilgus, Eklund, Maggin, Taylor, and Allen (2018) regarding the Student Risk Screening Scale (SRSS) literature provides an illustration of this phenomenon. Results indicated that although the SRSS correlated on average at .71 and .67 with broad behavior and externalizing rating scales, respectively, correlations with extant externalizing data (e.g., ODRs) and academic data (e.g., CBM scores) were equal to .41 and .46, respectively. Although such differences are to be expected and may be attributable to shared method variance, they nevertheless signal a concern. Screening is most useful when it is capable of predicting outcomes that are meaningful to schools. Although a screening measure's correlation with a criterion rating scale may be indicative of the former's construct validity, correlations with ecologically valid outcomes speak to a measure's capacity to provide information of paramount concern to schools.

Second, despite its importance and critical nature within MTSS frameworks, teachers can be somewhat unaccepting of mental and behavioral health practices, including universal screening. This limited acceptance might be the result of multiple factors. For instance, teachers might feel that they do not have the capacity or expertise to act on screening results. Most teachers enter the profession with little to no training in identifying and addressing student mental and behavioral health concerns (Koller & Bertel, 2006). Reinke, Stormont, Herman, Puri, and Goel (2011) found that although 89% of teachers believed that schools should address student mental health concerns, only 34% of teachers felt they had the skills necessary to support these needs. Furthermore, teachers can have misperceptions regarding the very nature of "mental health risk," considering it to be defined by the presence of problem behaviors alone without considering the extent to which students are also engaging in adaptive behaviors (Graham, Phelps, Maddison, & Fitzgerald, 2011; Koller & Bertel, 2006).

Limited acceptability of screening procedures can also be related to concerns regarding the feasibility of screening and the services that will result from student identification. A recent survey of MTSS schools found that, of those not engaging in mental and behavioral health screening, approximately 35% reported their lack of screening was due to limited budgetary funds (Bruhn et al., 2014). An additional 15% expressed concerns regarding their capacity to provide services to identified students. Schools are understandably predominantly focused on ensuring the academic success of their students. Although mental and behavioral health certainly relate to

academic functioning, the indirect nature of the relationship between these areas suggests that mental and behavioral health will always be a secondary focus, particularly within the context of limited educator time and school resources. Accordingly, it is necessary that universal screening be highly efficient, low cost, and communicated to teachers as such. That is, if screening is not perceived as being sufficiently feasible, educators may still hesitate to adopt the practice.

### Training Teachers to Complete Assessments

The importance of addressing the screening barriers should not be understated. Universal screeners will not be useful to schools if they are not strong predictors of relevant outcomes. Furthermore, the value of screening data could be further lessened if it is not collected systematically or with rigor. Although an administrator could require screening data collection, if teachers do not perceive the practice and resulting service delivery as feasible, they might not dedicate adequate attention and effort to the screening process. This could ultimately result in the collection of screening data with questionable validity and utility.

One means by which to promote the predictive utility and acceptability of universal screening is systematic teacher training. A line of research has recently emerged regarding the training of teachers to complete student ratings. Multiple studies have suggested that the use of a brief training protocol, which includes opportunities to practice rating with performance feedback, can result in more accurate ratings of student behavior, as well as the functions of that behavior (Chafouleas, Riley-Tillman, Jaffery, Miller, & Harrison, 2015; Kilgus, Kazmerski, Taylor, & von der Embse, 2017). This research also suggested that teachers who received systematic training are more accepting of behavior assessment methods relative to those who have not received such training (LeBel, Kilgus, Briesch, & Chafouleas, 2010).

To date, researchers have yet to consider the influence of systematic training on universal screening for mental health concerns. Within applied settings, screening implementation commonly reflects that of traditional rating scales, where it is presumed unnecessary to provide users with specific training prior to the completion of ratings. However, such presumptions have existed in the absence of research, suggesting an opportunity to examine the potential benefit of such training procedures. In the context of the previously described training literature, it is hypothesized that systematically training teachers to complete ratings and better understand the importance of their role in identification and referral of students with mental health concerns may promote the acceptability and predictive utility of universal screening (Atkinson & Hornby, 2002; Rothi, Leavey, & Best, 2008).

### Purpose of This Study

Effective implementation and use of MTSS frameworks and integration of mental health services rely on teachers who are thoroughly trained and supported (Durlak,

Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Better equipping teachers to identify risk and promote well-being is an essential element of these processes (Han & Weiss, 2005). Understanding the influence of teacher training on construct knowledge and use of assessment to identify behavioral and mental health risk may provide important insights into how to support overall decision making at Tier 1. The goal of the present investigation was to evaluate the influence of a pilot teacher training program that addresses teacher knowledge of mental and behavioral risk and efficacy of universal screening with regard to the validity of universal screening data in predicting student end-of-year outcomes. The research questions are threefold. First, will trained teachers report greater levels of screening acceptance and stronger beliefs that teachers should engage in behavioral and mental health assessment relative to untrained teachers? We hypothesize that trained teachers will report higher levels of acceptability, feasibility, and independence of screening implementation than untrained teachers. Second, to what degree are Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) scores correlated with academic and behavioral outcomes across time, and are these relationships different among trained and untrained teachers? It was hypothesized that SAEBRS scores would moderately correlate with end-of-year disciplinary infractions and absences, as well as academic outcomes (math, comprehension, literature scores). It was further hypothesized that the relationship between SAEBRS scores and academic and behavioral outcomes would be stronger within the trained teacher group. Third, to what degree does teacher-reported acceptability, understanding of constructs, and feasibility of screening influence the relationship between SAEBRS scores and academic and behavioral outcomes? It was hypothesized that higher perceptions of assessment usability would, in turn, lead to more significant relationships between SAEBRS scores and academic and behavioral outcomes.

## METHOD

The current study included four elementary schools in an urban center in the northeastern United States. A total of 91 teachers (57 treatment, 34 control) screened 1,158 students. Teacher participants were primarily female (89%) and reported ethnic/racial backgrounds to be White (59%), Black or African American (27%), Hispanic (8.7%), and Asian (3.3%). Most of the teachers had between 1 and 5 years of experience (51.1%), followed by 5 to 10 years (27.2%), less than 1 year (10.9%), 10 to 15 years (7.6%), and more than 15 years (3.3%). The average class size was 27 students ( $SD = 13.04$ ). There were 228 kindergarten students, 197 in first grade, 190 in second grade, 207 in third grade, 212 in fourth grade, and 124 in fifth grade. The student gender breakdown was 51.1% male and 48.9% female. Students were primarily Black/African American (77.6%), followed by Hispanic (16.1%), White (4.1%), Asian (0.7%), and multiracial (0.9%).

There were several measures used in the present investigation. The Social, Academic, and Emotional Behavior Risk Screener–Teacher Rating Scale (SAEBRS-TRS; Kilgus & von der Embse, 2014) is a 19-item, Likert-type teacher-report scale of behavioral risk that includes a total behavior risk score, as well as three subscales: Social Behavior (6 items,  $\alpha = .90$  within the current sample), Academic Behavior (6 items,  $\alpha = .91$ ), and Emotional Behavior (7 items,  $\alpha = .88$ ). Higher scores on the SAEBRS are indicative of better student functioning. The SAEBRS-TRS is one form of a broader multi-informant assessment suite that is available commercially ([www.fastbridge.org](http://www.fastbridge.org)) and includes a parent form (SAEBRS-PRS) and student self-rating (SAEBRS-SRS; von der Embse, Iaccarino, Mankin, Kilgus, & Magen, 2017; von der Embse, Kilgus, Iaccarino, & Levi-Nielsen, 2017). The SAEBRS-TRS has exhibited strong psychometric properties across multiple studies (Eklund, Kilgus, von der Embse, Beardmore, & Tanner, 2016; Kilgus, Eklund, von der Embse, Taylor, & Sims, 2016; Kilgus, Sims, von der Embse, & Taylor, 2016; von der Embse, Pendergast, Kilgus, & Eklund, 2016).

The Usage Rating Profile-Assessment (URP-A; Chafouleas, Miller, Briesch, Neugebauer, & Riley-Tillman, 2012) is a 28-item instrument that assesses factors that may influence adoption and use of school-based assessments and utilizes a 6-point Likert-type scale ranging from strongly disagree to strongly agree. The four subscales of primary interest in the present study included acceptability (e.g., “assessment is an effective choice for understanding a variety of problems”), understanding (e.g., “I understand how to use this instrument”), feasibility (e.g., “this assessment is too complex to be carried out accurately” [reverse scored]), and system supports (e.g., “I would need consultative support to implement this assessment”). Lower scores on the system support subscale indicate greater independence in assessment implementation. The reported URP-A internal consistency across subscales ranged from .63 to .90 (Miller, Chafouleas, Riley-Tillman, & Fabiano, 2014).

The Mental Health Needs and Practices Survey (MHNPS; Reinke et al., 2011) is a 44-item, Likert-type scale that measures participant attitudes on the role of schools in promoting mental health and perceptions and knowledge of evidence-based mental health treatment in schools. The survey also addresses mental health concerns (e.g., whether the teacher has taught a student with depression), current training in mental health practices, and perceived barriers to services. The MHNPS has demonstrated acceptable internal consistency across each subscale ( $\alpha = .78$  to  $.86$ ; Reinke et al., 2011).

Lastly, all academic and behavioral outcome data across grades were aggregated at the end of the second and third quarters (i.e., January and April) of the academic year. At the time of this study, the fourth-quarter data were not yet complete. Academic data (reported on a 0–100 scale) included composite scores in the domains of literature/reading, comprehension, and math. In addition, a 0–6 score was assigned for the Fountas-Pinnell (FP; Fountas & Pinnell, 2007) reading

level and a school-specific measure of writing. Behavioral outcomes included the number of in-school suspensions (ISS), out-of-school suspensions (OSS), absences, and disciplinary infractions at Level 1 (minor) and Level 2 (major).

The development of teacher training materials occurred over several stages. Research has supported the need for and promise of training teachers to promote student mental health (Han & Weiss, 2005; Jorm, Kitchener, Sawyer, Scales, & Cvetkovski, 2010). Barriers to this goal include (a) a lack of training at the preservice or university level (Franklin et al., 2012) and (b) existing programs with limited feasibility (e.g., prohibitive cost, extensive time) and evidence of effectiveness in a school setting (Jorm et al., 2010). In response, a novel 90-min pilot teacher training program was developed for the present study featuring four primary content domains, including (a) mental and behavioral health symptomology in schools; (b) methods to identify risk (e.g., universal screening, teacher referral) and the teacher role in these processes; (c) a review of SAEBRS items, including video examples, in a classroom setting; and (d) practice ratings based upon ambiguous video examples with performance feedback (Joyce & Showers, 2002; Rose & Church, 1998).

The training was developed in two stages. First, advanced graduate students and the first author conducted an extensive review of existing evidence-based training materials that specifically addressed mental and behavioral health issues (e.g., Youth-Mental Health First Aid; Jorm et al., 2010). Themes were identified, reviewed, and then adapted for a school setting. For example, common myths (e.g., comorbidity of disorders) and modern conceptualizations of mental health (e.g., Dual-Factor Model; Suldo & Shaffer, 2008) as displayed in school were included. The content was reviewed by five experts in the mental health field to determine appropriateness for inclusion within the training.

Second, approximately 70 short video clips (between 5–15 s) were filmed using child actors that displayed school-based examples of the 19 items on the SAEBRS. Each SAEBRS item had at least three corresponding video clip examples. Trained graduate students independently reviewed each clip to determine alignment to the 19 items. Consensus was reached through review of ratings and subsequent discussion to select one corresponding video example for each item. In addition, four video clips were identified for use in the performance feedback portion of the training. For each of the four performance feedback clips, a correct behavior rating was determined by a 100% consensus across five raters, and any clips with less than 100% consensus were not considered for inclusion.

Participating elementary schools were matched based upon similar student demographic characteristics. Schools were drawn from two distinct neighborhoods in a large metropolitan area and were then randomly assigned to a treatment or control condition such that one school from each neighborhood received the training and the other did not. Researchers scheduled the training during a professional development day in the fall of 2016. A 90-min training was

then delivered to the two elementary schools in the treatment group. A primary goal of the training was to promote teacher construct and parameter knowledge (e.g., what is and is not risk; Evans, Allen, Moore, & Strauss, 2005) and to emphasize the teachers' role in facilitating prevention and early intervention (Reinke et al., 2011). Training fidelity was assessed via self-report and was deemed to be 100% relative to the training checklist. The presenter (i.e., researcher) discussed content knowledge and teachers' role and provided video examples of SAEBRS items. The video clips included text descriptions of the items with examples and nonexamples, as well as a voiceover that described to the viewer who and what they were watching. Next, the presenter played four video clips of student behavior and asked the participating teachers to rate which item on the SAEBRS best represented each clip. Responses were reviewed with the teachers, and a discussion followed to address teacher rating questions. Teachers in both schools reported high levels of accuracy (averages of 86% and 82% across clips, respectively) on the four video clips.

Lastly, teachers completed a counterbalanced SAEBRS and a teacher nomination checklist for each student in their respective classrooms, as well as the URP-A and a measure of teacher's willingness and ability to engage in universal screening (Reinke et al., 2011). Teachers in the control condition were provided a one-page handout that described an overview of universal screening and directions on how to complete the assessment. All teachers across both conditions completed the assessments within the same 2-week period.

Upon completion of all measures, a multistage data analytic plan was employed within the present study. Using the recommendations from Tabachnick and Fidell (2013), study variables were assessed for assumptions of normality and multicollinearity. Most variables were acceptably normal, with skewness falling below  $\pm 2$  and kurtosis indices falling below  $\pm 7$  (Curran, West, & Finch, 1996). However, ISS and OSS were severely skewed, and the Level 1 and Level 2 disciplinary infractions exhibited slight skew and kurtosis. As such, variables were adjusted with a logarithmic transformation (Tabachnick & Fidell, 2013). Both OSS and ISS continued to fall slightly outside of the range of approximate normality and should be interpreted with caution. The bootstrapping technique within the moderation analysis allowed for the inclusion of skewed variables (Russell & Dean, 2000).

Missing data were addressed with different methods for each research question. With regard to teacher responses (i.e., Research Question 1) and the moderation analyses (i.e., Research Question 3), listwise deletion was employed. Chi-square difference tests indicated no significant differences among those teachers who did not complete the full survey. All students had FP reading scores, the primary academic outcome for students in kindergarten, first, and second grades. Full information maximum likelihood was utilized to address missing data for the modeling approach used within Mplus for Research Question 2.

To address the first research question, SPSS Version 24.0 was used to conduct an independent samples *t* test to

determine teacher reported differences on the URP-A subscales (acceptability, understanding, feasibility, system support) and several items on the MHNPS that explicitly pertain to teacher role in assessment (e.g., “teachers should screen for mental health problems,” “teachers should conduct behavioral assessments”) and reasons why children with mental health needs do not receive services (e.g., “lack of staff coaching and training in mental and behavioral health,” “lack of ongoing monitoring of students with mental health needs,” “difficulty identifying children with mental health needs”). Levene’s test for equality of variances was used to examine the homogeneity of variance assumption, and the Welch–Satterthwaite method was used for cases with unequal variances.

To address the second research question, bivariate correlations and a multigroup path analysis were used to evaluate the relationship between SAEBRS scores and academic/behavioral outcomes between trained and untrained teachers. First, a series of Pearson product-moment correlation ( $r$ ) coefficients were calculated in examining the relation between each SAEBRS scale and the various behavioral and academic outcome variables. Correlations were calculated within each group (trained and untrained). A more stringent critical  $p$  value of .01 was used given the large number of correlations considered. Second, Mplus 7.11 was used to conduct a path analysis using a multigroup structural equation model (MG-SEM). The full information maximum likelihood estimator, which is robust against violations of normality, was used to (a) address missing data, (b) test path models, (c) obtain estimates of model parameters, and (d) produce goodness-of-fit indices (Peters & Enders, 2002). The following indices were considered in evaluation model fit: comparative fit index (CFI), root-mean-square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Using Hu and Bentler’s (1999) guidelines, the following levels were considered to be acceptable model fit:  $RMSEA \leq .08$ ,  $CFI \geq .95$ , and  $SRMR \leq .05$ .

Academic and behavioral models were run separately and in accordance with prior SAEBRS theory and empirical evidence for predictive validity (see Kilgus, Eklund, et al., 2016; Kilgus, Sims, et al., 2016; von der Embse et al., 2016). That is, SAEBRS total behavior score was the primary predictor of behavioral outcomes (OSS, ISS, disciplinary Levels 1 and 2) and the SAEBRS Academic Behavior subscale was the primary predictor of academic outcomes (math, literature, comprehension, and FP). The CLUSTER command was utilized to address nesting of students within classrooms. Consistent with recommendations (Pendegast, von der Embse, Kilgus, & Eklund, 2017), the MG-SEM was tested for cross-group invariance to determine if the model with the grouping variable (trained or untrained) was significantly different from the initial unconstrained model. The Satorra-Bentler scaled chi-square and change in CFI and RMSEA were used to examine differences among nested models. First, configural invariance (i.e., the same indicators are loading on the same factors across groups) was examined by running separate models for trained and untrained teachers. Then, metric

invariance (i.e., the magnitudes of the relationships among variables/paths are similar) was assessed (Meredith & Teresi, 2006) by constraining the paths from predictor to outcome to be equal between groups. If the metric model was invariant, then there would be no significant differences between groups in the strength of paths between predictor and outcome. Third, for models that were demonstrated to be noninvariant, comparisons were made between trained and untrained groups.

To address the third research question, the PROCESS (Hayes, 2013) plugin within SPSS Version 24.0 was used to examine a series of moderation models between the potential influence of teacher reported acceptability, understanding, feasibility, and system support and the previously identified significant pathways between SAEBRS predictor and academic and behavioral outcomes. PROCESS is a bootstrapping procedure that creates a representation of the sampling distribution with a preset number of iterations (5,000 in the present investigation). Per recommendations from Hayes (2013), significant model change in  $R^2$  with the addition of the interaction term (e.g., Acceptability  $\times$  SAEBRS Total Score) was necessary before probing simple slope estimates. Predictors were mean-centered to facilitate interpretation, and simple slopes were produced at  $\pm 1$  standard deviation from the interaction mean (i.e., “low” or “high” in Table 1). As with correlational analyses, given the larger number of moderation analyses, a more stringent critical  $p$  value of  $< .01$  was used in evaluating statistical significance.

## RESULTS

Data analyses were conducted to examine the extent to which trained teachers demonstrated greater levels of screening acceptability and attitudes toward their role in screening than untrained teachers. Our review of the Levene’s test indicated that several of the variables violated the homogeneity of variance assumption. Thus, the Welch–Satterthwaite method was utilized and “equal variances not assumed” was used for significance testing. On the URP-A subscales, teachers in the training group reported significantly higher levels of acceptability,  $t(26.26) = 2.34$ ,  $p = .027$ ; understanding,  $t(27.19) = 2.34$ ,  $p = .027$ ; and feasibility,  $t(29.39) = 2.89$ ,  $p = .007$ , and significantly lower scores on system support (suggesting greater levels of independence in implementation),  $t(34.42) = -2.42$ ,  $p = .021$ , than did the untrained teachers with regard to SAEBRS. Trained teachers also reported significantly higher agreement that teachers should screen for mental health concerns,  $t(55.14) = 3.25$ ,  $p = .002$ , and conduct behavioral assessments,  $t(39.36) = 2.30$ ,  $p = .027$ . In addition, trained teachers reported stronger beliefs than untrained teachers that a lack of ongoing monitoring for students with mental health concerns may be a reason that students do not receive services,  $t(52.10) = 2.11$ ,  $p = .039$ . Despite higher ratings, there were no significant differences identified by the lack of staff coaching and difficulty identifying students that could explain why students may not receive services.

**Table 1. Simple Slope Estimates From the Moderated Models**

Predictor	Outcome	Moderator (Level)	<i>B</i>	<i>SE</i>	<i>p</i>
<b>SAEBRSTot</b>					
	DiscLv1	Accept (low)	-0.23	0.03	< .001
	DiscLv1	Accept (high)	-0.53	0.05	< .001
	DiscLv1	Understand (low)	-0.27	0.03	< .001
	DiscLv1	Understand (high)	-0.51	0.05	< .001
	DiscLv1	Feasibility (low)	-0.21	0.03	< .001
	DiscLv1	Feasibility (high)	-0.54	0.05	< .001
	DiscLv2	Understand (low)	-0.15	0.02	< .001
	DiscLv2	Understand (high)	-0.26	0.03	< .001
<b>SAEBRSAca</b>					
	MathQ3	Accept (low)	2.44	0.35	< .001
	MathQ3	Accept (high)	1.15	0.30	< .001
	MathQ3	Understand (low)	2.50	0.36	< .001
	MathQ3	Understand (high)	1.29	0.30	< .001
	MathQ3	Feasibility (low)	2.38	0.37	< .001
	MathQ3	Feasibility (high)	1.14	0.30	< .001
	LitQ3	Understand (low)	2.02	0.37	< .001
	LitQ3	Understand (high)	0.95	0.31	< .001
	LitQ3	Feasibility (low)	2.19	0.37	< .001
	CompQ3	Understand (low)	2.67	0.30	< .001
	CompQ3	Understand (high)	1.39	0.35	< .001
	CompQ3	Feasibility (low)	2.72	0.31	< .001
	CompQ3	Feasibility (high)	1.41	0.38	< .001

*Note.* Simple slope estimates were only included if the interaction term was significant. There were no significant interactions within the behavior outcomes for out-of-school suspensions, in-school suspensions, and absences or in the academic outcomes for FP reading level and writing. Academic models only included Quarter 3 outcomes. SAEBRSTot = Social, Academic, and Emotional Behavior Risk Screener total behavior score; DiscLv1 = Discipline Level 1; DiscLv2 = Discipline Level 2; SAEBRSAca = Social, Academic, and Emotional Behavior Risk Screener Academic Behavior score; MathQ3 = Quarter 3 math score; LitQ3 = Quarter 3 literature score; CompQ3 = Quarter 3 comprehension score.

Second, analyses were used to examine the relationship between SAEBRS scores and academic and behavioral outcomes and whether these relationships were different between trained and untrained teachers. See Table 1 for a review of correlational findings within trained and untrained groups. In accordance with hypotheses, SAEBRS scales were more highly associated with behavioral outcomes relative to academic outcomes. The exception to this was the SAEBRS Academic Behavior scale, which was consistently and statistically significantly correlated with academic outcomes. In comparing correlations across groups, the majority of correlational values were found to be quite similar in terms of magnitude. Three exceptions to this trend were noted. First, correlations between SAEBRS scales and FP scores were higher within the trained group. Specifically, the average absolute correlational value within the trained group was

equal to .20, compared to .09 within the untrained group. Second, trained group correlations were also found to be higher for SAEBRS scales and OSSs. The average absolute correlational value within the trained group was equal to .32, whereas the average untrained group value was equal to .19. Third, correlations between the SAEBRS Social Behavior scale and academic benchmark scores tended to be lower within the trained group. Specifically, the average absolute correlational value within the trained group was equal to .14, compared to .05 within the untrained group.

Second, a series of MG-SEMs were examined to determine differences between trained and untrained teachers on SAEBRS predictors and academic (i.e., academic model) and behavioral (i.e., behavior model) outcomes. The null (unconstrained) academic model, which included the SAEBRS Academic subscale and the second and third quarters of math,

literature, comprehension, and FP scores, exhibited good fit to the data (RMSEA = .032, CFI = .998, SRMR = .080). Next, separate models were examined and found to exhibit strong fit to the data for trained (RMSEA = .001, CFI = .998, SRMR = .060) and untrained teachers (RMSEA = .047, CFI = .996, SRMR = .092). To examine metric invariance (i.e., equality of path coefficients across groups), a final constrained model was examined and found to have strong fit to the data (RMSEA = .033, CFI = .996, SRMR = .089). The Satorra-Bentler chi-square test indicated a nonsignificant difference,  $\chi^2(8) = 13.51, p = .096$ , and changes in CFI ( $\leq .01$ ) and RMSEA ( $\leq .015$ ) were negligible (Chen, 2007). As such, metric invariance was established, and there were no significant path differences between trained and untrained teacher groups.

The null behavior model was then analyzed and found to exhibit good fit to the data (RMSEA = .051, CFI = .993, SRMR = .010). In an approach similar to that employed for the academic modeling strategy, separate models were examined and found to demonstrate good fit for the trained group (RMSEA = .032, CFI = .997, SRMR = .014) and untrained group (RMSEA = .071, CFI = .988, SRMR = .019). The fully constrained behavior model (i.e., paths constrained to be equal across groups) had adequate fit to the data (RMSEA = .079, CFI = .968, SRMR = .053). In comparing the null and fully constrained models, the Satorra-Bentler chi-square test

indicated a significant difference,  $\chi^2(4) = 25.42, p < .001$ , and there were differences in CFI and RMSEA greater than suggested cutoffs ( $\leq .01$  and  $\leq .015$ , respectively), suggesting metric noninvariance and that the models between trained and untrained teacher groups differed. Lastly, paths were compared between SAEBRS predictor and behavior outcomes. Paths were largely similar between groups for disciplinary levels (e.g.,  $b = -.590$  [untrained] versus  $b = -.594$  [trained]) and absences (e.g.,  $b = -.201$  [untrained] versus  $b = -.167$  [trained]). However, the trained teacher group demonstrated a stronger relationship between SAEBRS total and OSS ( $b = -.406$ ) than did the untrained group ( $b = -.258$ ).

Finally, a moderation analysis was employed to examine whether teacher-reported acceptability, construct understanding, and feasibility of screening influenced the relationship between SAEBRS scores and related academic and behavioral outcomes. A total of 27 (four behavior outcomes  $\times$  three distinct interactions + five academic outcomes  $\times$  three distinct interactions) moderation models were examined. A summary of simple slope estimates from models with a significant interaction term is presented in Table 2. There were no significant moderators of the relationship between SAEBRS total score and OSS, ISS, and absences. In contrast, acceptability ( $b = -.02, SE = .01, p < .001$ ), understanding ( $b = -.03, SE = .01, p < .001$ ), and feasibility ( $b = -.02, SE = .01, p < .001$ ) were found to moderate the relationship

**Table 2. Correlations Among Outcome Variables**

Outcome	Untrained Group				Trained Group			
	Total	Social	Academic	Emotional	Total	Social	Academic	Emotional
In-School Suspensions	-.04	-.04	.02	-.04	-.06	-.06	-.07	.01
Out-of-School Suspensions	*-.21	*-.29	*-.12	*-.14	*-.35	*-.38	*-.22	*-.32
Absences	*-.20	*-.16	*-.18	*-.18	*-.17	-.08	*-.22	*-.13
Minor Infractions	*-.49	*-.58	*-.32	*-.34	*-.56	*-.61	*-.40	*-.42
Major Infractions	*-.51	*-.60	*-.32	*-.39	*-.44	*-.50	*-.32	*-.33
Fountas & Pinnell Q2	.00	-.03	*.15	*-.14	*.23	*.16	*.27	*.16
Fountas & Pinnell Q3	-.01	-.08	*.15	*-.13	*.23	*.14	*.27	*.16
Writing Q2	.02	.02	*.14	*-.14	.10	.03	*.15	.07
Writing Q3	.04	-.04	*.18	-.04	.02	-.05	.13	-.04
Math Q2	*.30	*.17	*.42	*.16	*.20	.03	*.34	.17
Math Q3	*.37	*.22	*.45	*.24	*.22	.06	*.33	*.20
Reading Q2	*.24	*.18	*.41	-.03	*.22	.09	*.39	.13
Reading Q3	*.22	.16	*.34	.02	.08	-.03	*.26	.00
Comprehension Q2	*.25	.13	*.39	.08	.13	.01	*.24	.11
Comprehension Q3	*.31	.16	*.44	*.17	*.27	.12	*.38	*.22

Note. Q = quarter.

\* $p < .01$ .

between Level 1 disciplinary infractions and SAEBRS total scores. As acceptability, understanding, and feasibility perceptions of SAEBRS increased, there was a stronger negative relationship between SAEBRS total score (i.e., higher scores are indicative of better student functioning) and Level 1 disciplinary infractions. With Level 2 disciplinary infractions, understanding had a significant moderating influence, ( $b = -.02$ ,  $SE = .01$ ,  $p < .001$ ), whereas acceptability and feasibility were nonsignificant.

On the academic outcomes, there were several significant moderators identified across models. However, the moderators moved in a direction opposite that which had been predicted. See Table 2 for a summary of the simple slope estimates. Of note, acceptability ( $b = -.08$ ,  $SE = .03$ ,  $p = .01$ ), understanding ( $b = -.13$ ,  $SE = .05$ ,  $p = .01$ ), and feasibility ( $b = -.10$ ,  $SE = .04$ ,  $p = .01$ ) all significantly moderated the relationship between the SAEBRS Academic subscale and math Quarter 3 scores. That is, as acceptability, understanding, and feasibility increased, there was a weaker relationship between SAEBRS Academic and math scores. These same patterns were exhibited with both understanding and feasibility and their relationships to literature and comprehension scores.

## DISCUSSION

Despite repeated calls for research to examine factors that promote teacher recognition of mental health risk (Reinke et al., 2011) and a need for greater training and consultative support (Franklin et al., 2012), guidance for practitioners and school psychologists seeking to support teachers in this crucial area has yet to emerge. Schools are increasingly becoming the de facto providers of mental and behavioral health services (Oakes et al., 2014); however, many at-risk students do not receive the necessary services in a timely fashion (Eklund & Dowdy, 2014). Given their frequency of contact with students, teachers are uniquely positioned to bridge the gap between identification of risk and facilitation of early intervention services. Many teachers lack training and content knowledge of mental and behavioral health risks (Reinke et al., 2011) but are often called upon to refer students reactively via office disciplinary referrals that may lead to over- and underidentification problems. Alternatively, screening without knowledge of assessment use and key constructs may lead to incorrect decisions (Evans et al., 2005). A key goal of the present investigation was to inform future lines of research by examining a pilot training program that focused on (a) mental and behavioral health content knowledge, (b) teacher role in identifying at-risk students, and (c) opportunities to practice rating and receive performance feedback (Joyce & Showers, 2002). Teacher responses to training and the potential relationship to identifying at-risk students were of primary interest. As such, an improved understanding of the potential influences of training on universal screening may, in turn, inform and facilitate future decision-making processes in schools.

Critical to the success of program implementation in schools is the training and support provided (Ringeisen, Henderson, & Hoagwood, 2003). Results from the pilot training program highlighted differences in teacher responses to the acceptability and use of universal screening in schools. Trained teachers reported higher levels of acceptability, feasibility, independence of use, and understanding of universal screening. Utility ratings may be important predictors of sustained and consistent assessment to better inform Tier 1 decision making within MTSS frameworks (Cook, Volpe, & Livanis, 2010). Moreover, trained teachers reported stronger beliefs in the efficacy of screening and conducting behavioral assessments, while recognizing that a lack of ongoing monitoring may lead to situations in which some children are not receiving essential services. Helping teachers to recognize risk and the importance of their roles in these decisional processes may facilitate improved screening implementation and effectiveness of Tier 1 services (Lynn et al., 2003; Richards, Pavri, Golez, Canges, & Murphy, 2007).

There is a clear need for assessment research that considers both the interpretation and use of various tools (Kane, 2013). This study adds to the growing literature base supporting the applied use of universal screening and extends the work by considering the influence of malleable rater characteristics. For example, there were significant correlations between fall ratings of student risk and spring academic and behavioral outcomes, and these relationships were higher for teachers who received training. In addition, MG-SEM analyses indicated significant differences between trained and untrained teachers in these predictive relationships in the behavioral outcomes, specifically with OSSs. However, there were no differences noted across groups with academic outcomes. Thus, it is important to consider the potential positive contributions of training to the predictive relationships within the behavioral domain, while better communicating the correspondence of behavioral risk with long-term academic outcomes. Future assessment research on psychometric validity may consider the influence of input (i.e., rater training) and how a rating scale functions in applied settings to advance assessment utility.

For training to be considered an important part of the assessment process, data should support the relationship between training outcomes (e.g., rater perceptions of acceptability), ratings of student risk, and distal academic and behavioral indicators. To that end, the present investigation offered mixed evidence for the pilot teacher training program. Within the behavioral domain, moderation analyses suggested that higher levels of assessment understanding by teachers led to stronger correspondence between SAEBRS scores and two levels of disciplinary infractions. Ratings of feasibility and acceptability were also significant for Level 1 disciplinary infractions, suggesting that training outcomes may serve as a moderator to these relations. However, the opposite was found within the academic domain in that higher utility ratings (acceptability, feasibility, understanding) led to a weaker relationship between SAEBRS academic

ratings and academic outcomes. There was a difference in the variance accounted for by the moderated academic models and the behavior models ( $R^2 = .13-.21$  for academic versus  $R^2 = .31-.45$  for behavior); thus, there may be other variables to consider in the relationship between academic behavior risk and distal academic outcomes.

### Limitations

It is important to note that these are initial findings from a pilot study and additional research is needed to refine and improve the training processes to identify malleable teacher factors (e.g., efficacy, acceptability of assessment use) as well as critical content domains, including a greater emphasis on the link between mental and behavioral risk and academic performance. As such, there were several important limitations within the present investigation. First, although schools were matched based on student demographic characteristics and then randomly assigned to treatment (i.e., training), comparisons were made between trained and untrained teachers with different students. Future research should consider alternative raters (e.g., students, parents) and alternative rating formats (e.g., systematic direct observations) with the same student population to establish a comparative criterion. However, it is unclear how between-rater convergence could be a desired outcome of training. Second, the sample for this study was drawn from a large urban school district in the northeastern United States, and generalization of results may be limited. Third, there were a limited number of academic outcomes for certain student grades (restricted primarily to grades 3–5). In addition, listwise deletion was used (with Research Questions 1 and 3), which may have introduced bias if data were not missing completely at random. Larger and more robust samples would allow for complex analyses to better approximate the multiple influences on the hypothesized relationships (e.g., moderated mediation). Data should also be collected across time to determine the stability of both risk and the correspondence of said risk with important distal outcomes. Fourth, there were inconsistent findings, particularly in the academic domain, regarding the potential moderating influence of utility ratings. In addition, beta values for several of the moderation analyses were relatively low. Evidence of these relationships is considered preliminary and caution is urged when considering evaluating the influence of teacher training on the relationship of screening with distal academic and behavioral outcomes. Readers are encouraged to use caution when interpreting the preliminary findings from this pilot training program, and much additional research is needed to elucidate a more nuanced understanding of essential training processes.

### Implications for Practice

Data-based decision making and accountability is one of the key domains of the NASP Practice Model (NASP,

2010). Assessment for decision-making purposes remains a primary and essential role of school psychologists (Castillo, Curtis, & Gelley, 2012), yet the nature of assessment activities is changing as school psychologists are increasingly involved in assessment activities across all three tiers of service (Armistead & Smallwood, 2012). MTSS require evidence-based assessment processes to inform intervention decision making. Universal screening is a key component of Tier 1, and its success is largely dependent on collaboration between school psychologists and teachers to communicate concerns about students (Feinstein, Fielding, Udvari-Solner, & Joshi, 2009). That is, school psychologists often rely on teachers—through consultation, disciplinary referrals, or traditional rating scale formats—to identify students who are at risk and in need of early intervention. However, the manner in which rating scales are presented can influence accuracy (Evans et al., 2005), and teacher knowledge of mental and behavioral risk may be incorrectly assumed. The present investigation provided initial evidence for a pilot training program to assist in this collaboration and to improve universal screening processes.

Evidence for the effectiveness of teacher professional development is equivocal, with interventions that are primarily “sit and get,” or lecture-based in format, exhibiting low efficacy (Joyce & Showers, 2002). An important component of effective professional development is the opportunity to practice and receive specific feedback. The pilot teacher training was 90 min and included evidence-based content (see Jorm et al., 2010) on mental and behavioral health symptomology, integrated video examples of school-based behavioral risk with a diverse population, and provided opportunities for practice and performance feedback. Qualitative follow-up interviews with randomly selected teacher participants indicated a high degree of training acceptability and utility (e.g., “the training helped me to better understand mental health issues in my students” and “made the screening process make sense”). To support adoption and use, training materials and related resources are in an electronic format and freely available upon request. In addition, the training program was relatively brief when compared to other training programs (e.g., Youth-Mental Health First Aid) and could be integrated into existing professional development time. Future research will further evaluate the retention and maintenance of the teacher trainings to inform the need for the size and scope of additional intervention. It is incumbent upon school psychologists to critically consider the costs and benefits of training teachers to improve early identification of, and, ultimately, services for, at-risk students.

### REFERENCES

- Adelman, H. S., & Taylor, L. (2010). *Mental health in schools: Engaging learners, preventing problems, and improving schools*. Thousand Oaks, CA: Corwin Press.
- Armistead, R. J., & Smallwood, D. L. (2014). The National Association of School Psychologists model for comprehensive and integrated school psychological services. In P. L. Harrison & A. Thomas (Eds.), *Best*

- practices in school psychology: Data-based and collaborative decision making* (pp. 9–24). Bethesda, MD: National Association of School Psychologists.
- Atkinson, M., & Hornby, G. (2002). *Mental health handbook for schools*. London, England: Routledge Falmer.
- Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K–12 schools. *Education and Treatment of Children, 37*, 611–634. doi:10.1353/etc.2014.0039
- Chafouleas, S. M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2012). *Usage Rating Profile–Assessment (URP-A)*. Storrs, CT: University of Connecticut.
- Chafouleas, S. M., Riley-Tillman, T. C., Jaffery, R., Miller, F. G., & Harrison, S. E. (2015). Preliminary investigation of the impact of a web-based module on Direct Behavior Rating accuracy. *School Mental Health, 7*(2), 92–104.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504.
- Clayton, C. J., Ballif-Spanvill, B., & Hunsaker, M. D. (2001). Preventing violence and teaching peace: A review of promising and effective anti-violence, conflict-resolution, and peace programs for elementary school children. *Applied and Preventive Psychology, 10*, 1–35. doi:10.1016/S0962-1849(05)80030-7
- Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention, 35*(4), 197–205.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. doi:10.1037/1082-989X.1.1.16
- Curtis, M. J., Castillo, J. M., & Gelley, C. (2012). School psychology 2010: Demographics, employment, and the context for professional practices—Part 1. *Communiqué, 40*(7), 28.
- Dever, B. V., Raines, T. C., & Barclay, C. M. (2012). Chasing the unicorn: Practical implementation of universal screening for behavioral and emotional risk. *School Psychology Forum, 6*(4), 108–118.
- Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation, 11*, 193–221. doi:10.1207/s1532768XJEPC1102\_04
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432. doi:10.1111/j.1467-8624.2010.01564.x
- Eklund, K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health, 6*(1), 40–49. doi:10.1007/s12310-013-9109-1
- Eklund, K., Kilgus, S., von der Embse, N., Beardmore, M., & Tanner, N. (2016). Use of universal screening scores to predict distal academic and behavioral outcomes: A multilevel approach. *Psychological Assessment, 29*(5), 486–499. doi:10.1037/pas0000355
- Evans, S. W., Allen, J., Moore, S., & Strauss, V. (2005). Measuring symptoms and functioning of youth with ADHD in middle schools. *Journal of Abnormal Child Psychology, 33*(6), 695–706. doi:10.1007/s10802-005-7648-0
- Feinstein, N. R., Fielding, K., Udvari-Solner, A., & Joshi, S. V. (2009). The supporting alliance in child and adolescent treatment: Enhancing collaboration among therapists, parents, and teachers. *American Journal of Psychotherapy, 63*(4), 319–344.
- Fountas, I. C., & Pinnell, G. S. (2007). *Fountas and Pinnell: Benchmark Assessment System 1: Grades K-2, Levels A-N*. Portsmouth, NH: Heinemann.
- Franklin, C. G. S., Kim, J. S., Ryan, T. N., Kelly, M. S., & Montgomery, K. L. (2012). Teacher involvement in school mental health interventions: A systematic review. *Children and Youth Services Review, 34*(5), 973–982. doi:10.1016/j.childyouth.2012.01.027
- Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*(3), 137–148. doi:10.1080/00461528409529290
- Graham, A., Phelps, R., Maddison, C., & Fitzgerald, R. (2011). Supporting children's mental health in schools: Teacher views. *Teachers and Teaching, 17*(4), 479–496. doi:10.1080/13540602.2011.580525
- Greer, F. W., Wilson, B. S., DiStefano, C., & Liu, J. (2012). Considering social validity in the context of emotional and behavioral screening. *School Psychology Forum, 6*(4), 148–159.
- Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*(6), 665–679. doi:10.1007/s10802-005-7646-2
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Hoover, J. J., Baca, L., Wexler-Love, E., & Saenz, L. (2008). National implementation of response to intervention (RTI): Research summary. Retrieved from <https://www.nasdse.org/Portals/0/NationalImplementationofRTI-ResearchSummary.pdf>
- Hu, L. T., & 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), 1–55.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*, 582–600.
- Johnson, C., Eva, A. L., Johnson, L., & Walker, B. (2011). Don't turn away: Empowering teachers to support students' mental health. *The Clearing House, 84*(1), 9–14. doi:10.1080/00098655.2010.484441
- Jorm, A. F., Kitchener, B. A., Sawyer, M. G., Scales, H., & Cvetkovski, S. (2010). Mental health first aid training for high school teachers: A cluster randomized trial. *BMC Psychiatry, 10*, 51. doi:10.1186/1471-244X-10-51
- Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Kahlberg, J. R., Lane, K. L., Driscoll, S., & Wehby, J. (2011). Systematic screening for emotional and behavioral disorders at the high school level: A formidable and necessary task. *Remedial and Special Education, 32*, 506–520. doi:10.1177/0741932510362508
- Kamphaus, R. W. (2012). Screening for behavioral and emotional risk: Constructs and practicalities. *School Psychology Forum, 6*(4), 89–97.
- Kamphaus, R. W., DiStefano, C., Dowdy, E., Eklund, K., & Dunn, A. R. (2010). Determining the presence of a problem: Comparing two approaches for detecting youth behavioral risk. *School Psychology Review, 39*, 395–407.
- Kane, M. T. (2013). Validating the interpretations and uses of tests scores. *Journal of Educational Measurement, 50*(1), 1–73. doi:10.1111/jedm.12000
- Kilgus, S. P., Eklund, K., Maggin, D. M., Taylor, C. N., & Allen, A. N. (2018). The Student Risk Screening Scale: A reliability and validity generalization meta-analysis. *Journal of Emotional and Behavioral Disorders, 26*(3), 143–155. doi:10.1177/1063426617710207
- Kilgus, S. P., Eklund, K., von der Embse, N. P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology, 58*, 21–39. doi:10.1016/j.jsp.2016.07.001
- Kilgus, S. P., Kazmerski, J. S., Taylor, C. N., & von der Embse, N. P. (2017). Use of direct behavior ratings to collect functional assessment data. *School Psychology Quarterly, 32*, 240–253. doi:10.1037/spq0000156
- Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Taylor, C. N. (2016). Technical adequacy of the Social, Academic, and Emotional Behavior Risk Screener in an elementary sample. *Assessment for Effective Intervention, 42*(1), 46–59. doi:10.1177/1534508415623269
- Kilgus, S. P., & von der Embse, N. P. (2014). *The Social, Academic, and Emotional Behavior Risk Screener—Teacher Rating Scale*. Minneapolis, MN: FastBridge Learning.
- Koller, J. R., & Bertel, J. M. (2006). Responding to today's mental health needs of children, families, and schools: Revisiting the preservice training and preparation of school-based personnel. *Education and Treatment of Children, 29*, 197–217.

- LeBel, T. J., Kilgus, S. P., Briesch, A. M., & Chafouleas, S. M. (2010). The impact of training on the accuracy of teacher-completed direct behavior ratings (DBRs). *Journal of Positive Behavior Interventions, 12*, 55–63. doi:10.1177/1098300708325265
- Lloyd, J. W., Kauffman, J. M., Landrum, T. J., & Roe, D. L. (1991). Why do teachers refer pupils for special education? An analysis of referral records. *Exceptionality: A Special Education Journal, 2*(3), 115–126. doi:10.1080/09362839109524774
- Lynn, C. J., McKay, M. M., & Atkins, M. S. (2003). School social work: Meeting the mental health needs of students through collaboration with teachers. *Children & Schools, 25*(4), 197–209. doi:10.1093/cs/25.4.197
- McIntosh, K., Ty, S. V., & Miller, L. D. (2014). Effects of school-wide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions, 16*, 209–218. doi:10.1177/1098300713491980
- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11), S69–S77.
- Miller, F. G., Chafouleas, S. M., Riley-Tillman, T. C., & Fabiano, G. A. (2014). Teacher perceptions of the usability of school-based behavior assessments. *Behavioral Disorders, 39*, 201–210. doi:10.1177/019874291303900405
- Miller, F. G., Cohen, D., Chafouleas, S. M., Riley-Tillman, T. C., Welsh, M. E., & Fabiano, G. A. (2015). A comparison of measures to screen for social, emotional, and behavioral risk. *School Psychology Quarterly, 30*, 184–196. doi:10.1037/spq0000085
- National Association of School Psychologists. (2010). *Model for Comprehensive and Integrated School Psychological Services*. 1–12. Retrieved from: <https://www.nasponline.org/standards-and-certification/nasp-practice-model>
- Oakes, W. P., Lane, K. L., Cox, M. L., & Messenger, M. (2014). Logistics of behavior screenings: How and why do we conduct behavior screenings at our school? *Preventing School Failure: Alternative Education for Children and Youth, 58*(3), 159–170. doi:10.1080/1045988X.2014.895572
- O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.
- Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: A global public-health challenge. *The Lancet, 369*(9569), 1302–1313. doi:10.1016/S0140-6736(07)60368-7
- Pearcy, M. T., Clopton, J. R., & Pope, A. W. (1993). Influences on teacher referral of children to mental health services: Gender, severity, and internalizing versus externalizing problems. *Journal of Emotional and Behavioral Disorders, 1*(3), 165–169. doi:10.1177/106342669300100304
- Pendergast, L. L., von der Embse, N. P., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology, 60*, 65–82. doi:10.1016/j.jsp.2016.11.002
- Perou, R., Bitsko, R. H., Blumberg, S. J., Pastor, P., Ghandour, R. M., Gfroerer, J. C., ... Huang, L. N. (2013). Mental health surveillance among children—United States, 2005–2011. *MMWR Surveillance Summary, 62*(2), 1–35.
- Peters, C. L. O., & Enders, C. (2002). A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing, 11*(1), 81–95.
- Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26*, 1–13. doi:10.1037/a0022714
- Richards, C., Pavri, S., Golez, F., Canges, R., & Murphy, J. (2007). Response to intervention: Building the capacity of teachers to serve students with learning difficulties. *Issues in Teacher Education, 16*(2), 55–64.
- Ringeisen, H., Henderson, K., & Hoagwood, K. (2003). Context matters: Schools and the “research to practice gap” in children's mental health. *School Psychology Review, 32*, 3–16.
- Roeser, R. W., van der Wolf, K., & Strobel, K. R. (2001). On the relation between social-emotional and school functioning during early adolescence: Preliminary findings from Dutch and American samples. *Journal of School Psychology, 39*, 111–139. doi:10.1016/S0022-4405(01)00060-7
- Rose, D. J., & Church, R. J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education, 8*, 5–35. doi:10.1023/A:1022860606825
- Rothi, D. M., Leavey, G., & Best, R. (2008). On the front-line: Teachers as active observers of pupils' mental health. *Teaching and Teacher Education, 24*(5), 1217–1231. doi:10.1016/j.tate.2007.09.011
- Russell, C. J., & Dean, M. A. (2000). To log or not to log: Bootstrap as an alternative to the parametric estimation of moderation effects in the presence of skewed dependent variables. *Organizational Research Methods, 3*(2), 166–185.
- Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45*, 193–223. doi:10.1016/j.jsp.2006.11.003
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 34*(4), 317–342. doi:10.1023/A:1021320817372
- Suldo, S. M., Gormley, M. J., DuPaul, G. J., & Anderson-Butcher, D. (2014). The impact of school mental health on student and school-level academic outcomes: Current status of the research and future directions. *School Mental Health, 6*(2), 84–98. doi:10.1007/s12310-013-9116-2
- Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review, 37*, 52–68.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Vannest, K. J. (2012). Implementing interventions and progress monitoring subsequent to universal screening. *School Psychology Forum: Research in Practice, 6*, 119–136.
- von der Embse, N. P., Iaccarino, S., Mankin, A., Kilgus, S. P., & Magen, E. (2017). Development and validation of the Social, Academic, and Emotional Behavior Risk Screener—Student Rating Scale. *Assessment for Effective Intervention, 42*(3), 186–192. doi:10.1177/1534508416679410
- von der Embse, N. P., Kilgus, S. P., Iaccarino, S., & Levi-Nielsen, S. (2017). Screening for student mental health risk: Diagnostic accuracy, measurement invariance, and predictive validity of the Social, Academic, and Emotional Behavior Risk Screener—Student Rating Scale (SAEBRS-SRS). *School Mental Health, 9*(3), 1–11. doi:10.1007/s12310-017-9214-7
- von der Embse, N. P., Pendergast, L., Kilgus, S. P., & Eklund, K. R. (2016). Evaluating the applied use of a mental health screener: Structural validity of the Social, Academic, and Emotional Behavior Risk Screener. *Psychological Assessment, 28*(10), 1265–1275. doi:10.1037/pas0000253
- von der Embse, N. P., Rutherford, L., Mankin, A., & Jenkins, A. (2018). Implementation of a trauma-informed assessment to intervention model in a large urban school district. *School Mental Health*. doi:10.1007/s12310-018-9294-z

Date Received: July 8, 2017

Date Accepted: November 13, 2017

Associate Editor: Tanya Eckert ■

### AUTHOR BIOGRAPHICAL STATEMENTS

Nathaniel P. von der Embse, PhD, is an assistant professor of school psychology at the University of South Florida. Dr. von der Embse is a nationally certified school psychologist and serves as an associate editor for the *Journal of School Psychology*. His research interests include universal screening for behavioral and mental health, teacher stress and student test anxiety, and training educators in population-based mental health services. He received the 2018 Lightner Witmer Award for early career scholarship from Division 16 of the American Psychological Association.

Stephen P. Kilgus, PhD, is an associate professor of school psychology at the University of Wisconsin-Madison. His primary research interests pertain to social-emotional and behavioral assessment and intervention. Dr. Kilgus currently serves as an associate editor for the *Journal of School Psychology*. He received the Lightner Witmer Award for early career scholarship from Division 16 of the American Psychological Association.

Katie Eklund, PhD, is an assistant professor in the school psychology program at the University of Wisconsin-Madison. She is a nationally certified school psychologist and licensed psychologist. Dr. Eklund's research on school mental health includes universal screening and early intervention for mental health risk, advocacy and policy promotion of school psychologists as mental health providers, and school climate and student safety.

Ethan Ake and Shana Levi-Nielsen are doctoral students in the College of Education at Temple University.