Psychological Assessment
2015, Vol. 27, No. 1, 215–227

1040-3590/15/$12.00 http://dx.doi.org/10.1037/a0037787

Mental Health Screening in Kindergarten Youth: A Multistudy Examination of the Concurrent and Diagnostic Validity of the Impairment Rating Scale

Erin Girio-Herrera, Melissa R. Dvorsky, and Julie Sarno Owens
Ohio University

Using a multistudy approach, we examined the utility of the Impairment Rating Scale (IRS; Fabiano et al., 2006) as a screening tool for detecting kindergarten children who are at risk for social, emotional, academic, and behavioral problems. In Study 1 ($N = 568$), we evaluated the concurrent validity, discriminant validity, and diagnostic efficiency of the parent and teacher IRS test score inferences in relation to scores from the Behavioral Assessment System for Children (BASC-2; Reynolds & Kamphaus, 2004). In Study 2 ($N = 242$), we addressed limitations in Study 1 and evaluated the concurrent validity, discriminant validity, and diagnostic efficiency of the parent and teacher IRS test score inferences in relation to scores from BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), quarterly grades, kindergarten reading competency tests, and daily behavior outcomes on a classwide discipline system. Results indicate moderate to strong concurrent and diagnostic validity utility for the teacher IRS test score inferences and low to moderate concurrent and diagnostic validity utility for the parent IRS test score inferences. IRS scores of 3 or 4 may represent appropriate cutpoints for determining risk status in kindergarten youth depending on school districts' intended use of the tool for screening. Implications for future research and practice in universal school-based screening are discussed.

Keywords: universal screening, mental health, schools, kindergarten, Impairment Rating Scale

Mental health challenges experienced early in childhood tend to be stable and predictive of negative outcomes later in life (e.g., Lochman & The Conduct Problems Prevention Research Group [CPPRG], 1995). Studies document that both internalizing and externalizing problems detected in preschool were predictive of internalizing and behavior problems 7–8 years later (Ashford, Smit, van Lier, Cuipers, & Koot, 2008; Pihlakoski et al., 2006). Such mental health problems are risk factors for negative outcomes, such as absenteeism, suspensions, academic underachievement, school dropout, antisocial behavior, and substance abuse (Lochman & The CPPRG, 1995; Tremblay et al., 1992; Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005).

Although many children experience mental health problems (Roberts, Attkisson, & Rosenblatt, 1998), only a small percentage are identified by "frontline gatekeepers" such as teachers or pediatricians (Briggs-Gowan, Horwitz, Schwab-Stone, Leventhal, & Leaf, 2000). This problem of unidentified children warrants the expansion of proactive, universal screening programs (Levitt, Saka, Romanelli, & Hoagwood, 2007) to improve early identification and potentially prevent and alter negative trajectories toward costly and debilitating long-term outcomes (Hill, Lochman, Coie, Greenberg, & The CPPRG, 2004).

Benefits of Universal School-Based Mental Health Screening

One advantage of universal school-based screening is that this systematic process involves screening *all* children, thereby reducing the chance that students are "missed." In contrast, other approaches such as nomination or referral processes, through which teachers submit names of children based on their own individual threshold of concern, may inherently miss at-risk students as a function of different thresholds across teachers and arbitrary limits on the number of students who can be nominated (Lloyd, Kauffman, Landrum, & Roe, 1991). Indeed, universal screening, via structured teacher ratings, has been shown to detect students who were not identified using traditional referral methods (Eklund et al., 2009). Further, school staff may use the screening results to (a) streamline the identification of target behaviors for assessment or early intervention; (b) create a baseline from which future monitoring can be benchmarked; (c) monitor student response to intervention at an individualized level; and (d) determine outcomes at a programmatic level (e.g., impact of a prevention program) (Dowdy, Ritchey, & Kamphaus, 2010). Increased efficiency in

This article was published Online First September 15, 2014.

Erin Girio-Herrera, Melissa R. Dvorsky, and Julie Sarno Owens, Department of Psychology, Ohio University.

Melissa R. Dvorsky is now at the Department of Psychology, Virginia Commonwealth University.

During the preparation of this article, Julie Sarno Owens was supported by grants from the Department of Education, Institute of Educational Sciences. We wish to extend our gratitude to the kindergarteners, parents, teachers, and administrators of Eastern Local, Federal Hocking Local, Gallia County Local, Logan-Hocking, Meigs Local, Southern Local, Vinton County, and Wellston City School Districts who participated in the project.

Correspondence concerning this article should be addressed to Erin Girio-Herrera, Porter Hall 200, Department of Psychology, Ohio University, Athens, OH 45701. E-mail: girio@ohio.edu

these areas facilitates early service use, thereby reducing the likelihood of future problems (Albers, Glover, & Kratochwill, 2007).

Kindergarten entry represents an optimal time to screen children given that only 50% of children in the country attend preschool (U.S. Census Bureau, 2000), but nearly all attend kindergarten. Further, 60% of youth with developmental delays or mental health problems have not been detected prior to starting school (e.g., Halfon et al., 2004; King & Glascoe, 2003). Despite this, the vast majority of schools in the United States *do not* screen for such problems (Romer & McIntosh, 2005). By obtaining parent and teacher report of problems via universal screening at kindergarten entry, school personnel have the opportunity to make proactive decisions to maximize success for teachers (e.g., distribution of at-risk students across classrooms) and students (e.g., plans for monitoring or early intervention).

Limitations of Current Mental Health Screening in Schools

Although most schools have a screening process for vision, hearing, early literacy skills, and, more recently, obesity, only 2% of schools nationwide screen all students for emotional and behavioral difficulties (Romer & McIntosh, 2005). Similarly, screening techniques for social, emotional, and behavioral problems that are gated (e.g., Systematic Screening for Behavior Disorders [SSBD]; Walker & Severson, 1992) begin with teacher nominations and typically only allow a restricted number of nominations. With gated procedures, all children are not evaluated in the same structured and consistent manner, and at-risk children could go undetected.

Additionally, there is limited information regarding the role of various raters in screening. Some studies suggest that multiple informants provide unique information (Achenbach, McConaughy, & Howell, 1987; Feil, Severson, & Walker, 1995), whereas others suggest that teacher ratings alone may be sufficient (e.g., Kamphaus, DiStefano, Dowdy, Eklund, & Dunn, 2010; Reynolds & Kamphaus, 1992). Unfortunately, many screening measures such as Ages and Stages (ASQ:SE; Squires, Bricker, & Twombly, 2002), SSBD (Walker & Severson, 1992), Parents Evaluation of Developmental Stages (PEDS; Glascoe, 1999), Student Risk Screening Scale (SRSS; Drummond, 1993), and Pediatric Symptoms Checklist (PSC; Little, Murphy, Jellinek, Bishop, & Arnett, 1994) only offer the option for one rater, limiting what may be learned across caregivers and settings.

Further, many tools are narrowly focused on one specific type of problem (DiStefano & Kamphaus, 2007). For example, the SRSS (Drummond, 1993) and the Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999) are focused on social and/or externalizing behavior problems and do not provide the opportunity to identify children with other equally debilitating problems (e.g., depression, anxiety, adaptive problems). Similarly, Teen Screen (Shaffer et al., 2004) focuses primarily on depression and suicide, but not other externalizing mental health problems.

Of measures that screen for a broad array of problems, many use two different forms (one with norms up to preschool age and one with norms for elementary school-age children). This makes it difficult for schools to use one measure to obtain parent report of child problems at kindergarten registration (spring prior to elementary school) and teacher report (fall after entering elementary school). This problem applies to the Behavioral Assessment System for Children, second edition (BASC-2; Reynolds & Kamphaus, 2004) and the Preschool Behavior Checklist (McGuire & Richman, 1986). Thus, there is a critical need for measures with norms that adequately represent 4– to 6-year-olds through the duration of the kindergarten entry process.

Lastly, there has been less emphasis on the feasibility and acceptability of screening tools and the screening process, the infrastructure necessary to support the process, the availability of appropriate interventions following screening, and the acquisition of information that is useful to stakeholders (Glover & Albers, 2007; Levitt et al., 2007). Screeners that are lengthy may lead to unreliable data collection, as raters may become fatigued or irritated by the time demands (e.g., Kamphaus et al., 2007). Costs associated with purchasing and scoring measures may be prohibitive to some school districts. Furthermore, the level of training required to administer, score, and interpret screening results is often unaddressed. Examples of measures that involve costs for purchasing the measure and scoring systems include the ASQ:SE (Squires et al., 2002), the BASC-2 Behavioral and Emotional Screening System (BESS) (Kamphaus & Reynolds, 2007), and PEDS (Glascoe, 1999). Although the BESS is brief, reliable, and predictive of future behaviors (DiStefano & Kamphaus, 2007; Kamphaus et al., 2007; Kamphaus et al., 2010), some school districts may be deterred by the cost, as well as the graduate-level training required to interpret results. Examples of other measures that require training or advanced degrees for administration, scoring, or interpretation include the ECBI (Eyberg & Pincus, 1999) and PSC (Little et al., 1994). Lastly, some such as the Strengths and Difficulties Questionnaire (Goodman, 1997) require minimal cost and training of personnel but have time demands related to calculating scores into subscales.

Given the strengths and limitations of existing screening tools, there remains a critical need for a brief, psychometrically strong tool that screens all children for a wide variety of social, emotional, behavioral, and adaptive problems that is free and in the public domain, requires no advanced degree to administer and score, allows for use by multiple raters, and is capable of screening kindergarten-aged children throughout the registration process and across the kindergarten year. (Because a full review of measures is beyond the scope of this article, readers are referred to the review of over 95 mental health screening tools conducted by the Northern California Training Academy (http://humanservices.ucdavis.edu/Academy/pdf/104056-MentalHealthLR.pdf).

The Impairment Rating Scale

The Impairment Rating Scale (IRS; Fabiano et al., 2006) assesses parent (eight items) and teacher (six items) perceptions of child impairment in multiple domains (i.e., academic, social, behavioral, family) that are critical to healthy development and school success (see the Measures section below for details). The strong psychometric properties of the IRS, including high test–retest reliability, construct validity, and discriminant validity, have been documented in a preschool sample, multiple elementary school-age samples (Fabiano et al., 2006), a large high school-age sample (Evans et al., 2013), as well as in treatment outcome studies with elementary students (Fabiano et al., 2010; Owens, Murphy, Richerson, Girio, & Himawan, 2008). A cutoff score of

3 discriminated between clinical and nonclinical groups of children (Fabiano et al., 2006). The IRS also has several characteristics that make it feasible within schools (free, publically available, brief, multirater options, and no advanced training required to administer and score). The IRS has never been examined specifically as a screener. However, previous research supports its consideration as a screener given the strong psychometric properties of score inferences in samples of children ages 4–6 (Fabiano et al., 2006), ability to differentiate between children with and without a mental health disorder (Fabiano et al., 2006), and ability to detect change in response to school-based intervention (Fabiano et al., 2010; Owens et al., 2008).

The Current Studies

The overall goal of the current investigation was to examine the IRS as a screening tool for detecting kindergarten children who are at risk for social, emotional, academic, and behavioral problems. Specifically, in Study 1, we evaluated the concurrent validity for children entering kindergarten in relation to scores from the BASC-2 (Reynolds & Kamphaus, 2004) by examining within- and between-rater bivariate correlations (Aim 1). As a test of discriminant validity, we examined area under the curve (AUC) statistics produced by receiver operator characteristic (ROC) analyses (Aim 2) and examined diagnostic efficiency statistics for multiple cutoff scores to determine the level of impairment on the parent and teacher IRS that best differentiates typical and at-risk children, as determined by the parent and teacher BASC-2 (Aim 3).

Given that a cutoff of 3 or higher has been identified in previous research (Fabiano et al., 2006), this cutpoint (as well as a score of 2 and 4) was examined to determine which would be most useful when screening for a variety of problems (i.e., internalizing, externalizing, and adaptive problems as assessed by the BASC-2). The BASC-2 preschool parent and teacher rating forms were used to define risk status because of its strong psychometric properties; frequent use in schools; and broad examination of behavioral, social, emotional, and adaptive problems. Further, replicated psychometric data on BESS scores (Kamphaus et al., 2007) were not available at the time Study 1 was initiated. It was hypothesized that (a) concurrent and diagnostic validity of the IRS test score inferences would be acceptable, as evidenced by moderate to high correlations with BASC-2 scores and moderate to high AUC statistics, (b) within-rater concurrent validity scores (e.g., teacher IRS scores and teacher BASC-2 scores) would be stronger than between-rater concurrent validity scores (e.g., parent IRS scores and teacher BASC-2 scores), and (c) a score of 3 on the IRS would be the optimal cutoff for detecting risk status according to the BASC-2, given previous literature (Fabiano et al., 2006).

Study 2 was developed to address the limitations of Study 1 and to determine whether the findings obtained in Study 1 could be replicated. Specifically, the aims of Study 2 were to evaluate the concurrent validity and diagnostic validity of the IRS test score inferences in relation to BESS scores and academic (grade-point average, test scores) and behavioral indicators of kindergarten success (i.e., daily behavior data) by examining within- and between-rater bivariate correlations (Aim 1) and AUC statistics (Aim 2), as well as what level of impairment on the parent and teacher IRS (cutoff of 2, 3, or 4) best differentiates typical and at-risk children as determined by the BESS and these important kindergarten outcomes (Aim 3).

Study 1

Method

Participants. The parents of 693 kindergarteners from 18 schools (63% of all enrolled) consented to participate, along with their teachers. Of the 693, 125 participants were not included in the final sample due to the child was age 6 or older at screening (n = 76; outside of the age norms for preschool version of the BASC-2) or incomplete data for one of the informants (n = 49; at least one item missing on the IRS or enough items missing on the BASC-2 to prohibit the use of the scoring program). The final sample consisted of 568 children. Those excluded did not differ from those included with regard to gender, race, socioeconomic status, or caregiver age. Child participants were 46.8% male and 95.1% Caucasian (2.8% classified as other and less than 1% as African American, Hispanic, Asian, and American Indian/Alaskan Native) and with a mean age of 5.48 (SD = 0.32). All teachers (n = 56) were female and Caucasian. The following data represent the highest education achieved by mothers and fathers in this sample: less than high school completion: 10.4% for mothers, 13.7% for fathers; high school completion: 31.2% for mothers, 48.2% for fathers; associate's degree or some college: 40.4% for mothers, 26.5% for fathers; bachelor's degree: 9.5% for mothers, 5.6% for fathers; graduate degree: 6.5% for mothers, 2.6% for fathers. This indicates that the sample is characterized by families of middle and lower socioeconomic status and is representative of the communities in the Appalachian region of the state from which the sample was drawn (U.S. Census Bureau, 2006).

Measures.

Parent demographic questionnaire. Parents provided information about child and family characteristics including child and parent age, child race, maternal and paternal education and employment, family income, and insurance status.

The BASC-2. The parent and teacher preschool versions (ages 2–5) of the BASC-2 (Reynolds & Kamphaus, 2004) were used to assess child emotional and behavioral functioning. Age-based T scores were used to determine the risk status (see the Data Preparation section below). Consistent with previous studies involving classification using the BASC-2 (e.g., Dowdy & Kamphaus, 2007), the Externalizing Problems (Hyperactivity and Aggression scales), Internalizing Problems (Anxiety, Depression, and Somatization scales), and Adaptive Skills (Adaptability, Functional Communication, Social Skills, and Activities of Daily Living scales) were used to determine risk status. Parent and teacher versions have undergone rigorous psychometric evaluation with a large, nationally representative sample; reliability and validity scores are strong (Reynolds & Kamphaus, 2004).

BASC-2 scores represent the criterion measure in Study 1.

The IRS. The IRS (Fabiano et al., 2006) is the measure of focus for Studies 1 and 2 and assesses perceptions of child impairment in multiple domains and overall impairment. The parallel parent (eight items) and teacher (six items) versions assess impairment via one question about each of the following domains: relations with peers, relations with teachers (teacher version), relations with parents (parent version), relations with siblings

(parent version), academic progress, self-esteem, classroom functioning (teacher version), family functioning (parent version), and overall impairment. The parent version includes two additional items: one that inquires about impairment in sibling relations and one that asks whether the child has a best friend. However, for the purpose of this study, only parallel items across both raters were analyzed.

The IRS was designed to measure individual domains of functioning using a single-item-per-domain format (Fabiano et al., 2006). The rating scale can be found at http://ccf.fiu.edu/for-families/resources-for-parents/printable-information/ (see assessment instruments). Previous studies have shown strong psychometric properties (see below); thus, the single-item-per-domain format was retained in the current study. Each item begins with, "How does the child's problems affect . . ." Both parent and teacher informants place an "X" on a visual analogue scale to signify their perceptions of child functioning along a continuum of impairment. The scale shows two anchor points, one at each end using 0 (*Not a problem at all/Definitely does not need treatment or special services*) and 6 (*Extreme problem/Definitely needs treatment and special services*), with no other anchors in between. Following completion, a transparency with all numbers equidistantly spaced from 0 to 6 is used to determine on which of the seven points the "X" is marked. It can be completed in less than 5 min and does not require advanced training to administer or score. The validity of score inferences derived from the IRS have been documented in several ways. First, in an elementary sample, the IRS overall impairment item is highly correlated (r = .77) with scores on the Children's Global Assessment Scale (Fabiano et al., 2006). Second, there is evidence of convergent and divergent validity for the domain-specific items. For example, the IRS peer item is more highly correlated with teacher-rated aggression (r = .64) than is the IRS academic item (r = .42) (Fabiano et al., 2006). In a high school sample, the IRS academic item is more highly correlated with the Academic Competence factor on the Classroom Performance Survey (CPS) (r = .73) than any other IRS item (Brady, Evans, Berlin, Bunford, & Kern, 2012). In the same sample, the IRS student–teacher relationship item is more highly correlated with the Interpersonal Competence factor of the CPS (r = .70) than any other IRS item. Third, in two separate samples of children with attention-deficit/hyperactivity disorder and matched comparisons (one of preschool and kindergarten-aged children and the other children ages 5–12), a cutpoint of 3 or higher discriminated between clinical and nonclinical samples (Fabiano et al., 2006). Lastly, IRS scores have demonstrated sensitivity to change as a function of treatment (Fabiano et al., 2010; Owens et al., 2008).

With regard to reliability, moderate to high test–retest reliability has been established at 2 months (r = .66–.98 [teacher]; r = .82–.95 [parent]), 4 months (r = .57–.84 [teacher]; r = .76–.91 [parent]), 6 months (r = .64–.89 [teacher]; r = .60–.89 [parent]), and 1 year (r = .40–.67 [teacher]; r = .54–.76 [parent]). Lastly, in the current study, the Cronbach's alpha coefficient across all domains for Study 1 was .91 (parent) and .96 (teacher), suggesting that collectively, all items are measuring a similar construct (i.e., impairment).

Procedure. This study was approved by the Institutional Review Board (IRB). Ten school districts (26 elementary schools) within a 1-hr radius of the university were offered participation. No exclusion or inclusion criteria were used with schools or children. Thus, participation was based on interest and agreement of school administration. Six districts (18 schools) agreed to participate, one district refused (one school), and three districts (seven schools) did not respond prior to the project start date. Parents were consented at kindergarten registration and open house events held at the schools (August through October). Less than 1% of parents refused participation if they attended the scheduled registration or open house. Thus, schools with lower consent rates were a function of low attendance at open house. Parents completed the demographic questionnaire, BASC-2, and IRS prior to their departure from the school event and received $10 for participation. Teachers of participating children were consented during grade-level meetings 8–12 weeks after the start of school to allow adequate time to become familiar with children. Each teacher was provided with one BASC-2 and one IRS per consented child and was compensated $25. The maximum age norm (5 years 11 months) was used to calculate the BASC-2 for children who were below age 6 at the time of parent consent but turned 6 at the time of teacher ratings.

Data preparation.

Definition of at risk. Parent and teacher BASC-2 rating scales were scored separately to determine child risk status (either "typical" or "at risk"). Teacher-based BASC-2 risk status was coded as "at risk" if teacher ratings on the BASC-2 produced a T score of 60 or greater on either the Externalizing Problems or Internalizing Problems Composites or a T score of 40 or lower on the Adaptive Skills Composite. The same procedure was used to define "at risk" with the BASC-2 parent scores (i.e., parent-based BASC-2 risk status). The above cutpoints were based on the results from the BASC-2 manual (Reynolds & Kamphaus, 2004), which indicates T scores of 60–65 (35–40 on Adaptive) are considered "at risk," whereas T scores of 65 or higher (35 or lower) are within the clinical range. According to this criterion, 117 (20.6%) children were identified as "at risk" by teachers, and 160 (28.2%) were identified as "at risk" by parents. This rate of risk was expected given the socioeconomic and geographic characteristics of the population screened (see Girio-Herrera, Owens, & Langberg, 2013; U.S. Census Bureau, 2006).

The IRS. Level of risk on the six domains of the IRS was assessed separately for parent and teacher ratings. Given that the IRS has not previously been examined as a screening tool, diagnostic efficiency statistics for cutoff scores of 2, 3, and 4 were examined.

Results

Descriptive statistics for the parent- and teacher-rated IRS domains for the total sample and for the parent- and teacher-based at-risk and typical samples are presented in Table 1.

Aim 1: Concurrent validity. To assess concurrent validity, bivariate correlations between the scores on the parent and teacher IRS and scores on the parent and teacher BASC-2 composites scores were examined (see Table 2). The teacher-rated IRS scores yielded moderate to high associations with the teacher-rated BASC-2 Externalizing Composite scores (rs ranged from .48 to .72) and Adaptive Composite scores (rs ranged from −.48 to −.57). Teacher-rated IRS scores yielded low to moderate correlations with the teacher-rated Internalizing Composite scores (rs

Table 1
IRS Scores for Kindergartners by Parent- and Teacher-Based Risk Status According to the BASC-2 (Study 1)

IRS domain	Total sample (N = 568)				Parent-based at risk (n = 160)		Parent-based typical (n = 408)		Teacher-based at risk (n = 117)		Teacher-based typical (n = 451)	
	M	*SD*	Skewness (*SE*)	Kurtosis (*SE*)	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Parent IRS scores												
Peers relationship	0.42	1.02	2.98 (.10)	9.31 (.21)	0.97*	1.48	0.21a*	0.66	0.51	1.16	0.40	0.99
Parent–child relationship	0.44	1.06	2.84 (.10)	8.14 (.21)	0.99*	1.56	0.23a*	0.69	0.52	1.23	0.42	1.02
Academic performance	0.37	0.90	2.99 (.10)	9.82 (.21)	0.81*	1.29	0.20a*	0.63	0.50	1.02	0.33	0.87
Self-esteem	0.41	0.99	2.98 (.10)	9.62 (.21)	0.89*	1.37	0.23a*	0.71	0.44	0.98	0.41	0.99
Family functioning	0.43	1.04	3.05 (.10)	10.10 (.21)	0.93*	1.45	0.23a*	0.74	0.41	1.00	0.43	1.06
Overall impairment	0.44	1.11	3.10 (.10)	10.01 (.21)	1.03*	1.69	0.21a*	0.65	0.58	1.32	0.41	1.05
Teacher IRS scores												
Peers relationship	0.83	1.40	1.81 (.10)	2.51 (.21)	1.20*	1.70	0.69b*	1.24	2.52*	1.77	0.39c*	0.86
Teacher–child relationship	0.71	1.37	2.07 (.10)	3.52 (.21)	1.01*	2.03	0.60b*	1.26	2.19*	1.94	0.33c*	0.82
Academic performance	1.08	1.73	1.46 (.10)	.84 (.21)	1.57*	2.03	0.89b*	1.56	2.91*	2.08	0.61c*	1.26
Self-esteem	0.93	1.49	1.61 (.10)	1.62 (.21)	1.33*	1.73	0.77b*	1.35	2.59*	1.77	0.49c*	1.04
Classroom functioning	0.85	1.52	1.83 (.10)	2.47 (.21)	1.26*	1.79	0.69b*	1.37	2.67*	1.99	0.38c*	0.91
Overall impairment	1.00	1.63	1.60 (.10)	1.46 (.21)	1.43*	1.865	0.83b*	1.50	2.99*	1.86	0.48c*	1.08

Note. Children are considered at risk if they received a *T* score of 60 or higher on either the Externalizing Problems or Internalizing Problems Composites, or a score of 40 or lower on the Adaptive Skills Composite of the BASC-2 for that rater. IRS = Impairment Rating Scale; BASC-2 = Behavioral Assessment System for Children, second edition.
a* Effect sizes (Cohen's *d*; Cohen, 1988) for differences in Parent IRS scores between parent-based at-risk and typical groups ranged from .60 to .66. b* Effect sizes (Cohen's *d*; Cohen, 1988) for differences in Teacher IRS scores between parent-based at-risk and typical groups ranged from .24 to .38. c* Effect sizes (Cohen's *d*; Cohen, 1988) for differences in Teacher IRS scores between teacher-based at-risk and typical groups ranged from 1.25 to 1.65.
* $p < .01$ (significant differences between the at-risk and typical groups within rater).

ranged from .24 to .46). When examining across raters, there was a lack of association or low correlation for the teacher-rated IRS scores with all parent-rated composite scores on the BASC-2 (*r*s range from .02 to .31). The parent-rated IRS scores yielded moderate associations with parent-rated BASC-2 Externalizing Composite scores (*r*s range from .30 to .49), and low or no association with all other parent-rated BASC-2 scores (*r*s all lower than .40) and teacher-rated BASC-2 scores (*r*s all lower than .32). As expected, correlations were stronger for within-rater scores (parent IRS and parent BASC-2; teacher IRS and teacher BASC-2) than cross-rater scores (parent IRS and teacher BASC-2; teacher IRS and parent BASC-2; see Table 2).

Aim 2: Discriminant validity. To measure the extent to which scores on the IRS discriminated between at-risk and typical groups as determined by the BASC-2, the diagnostic validity of the parent and teacher IRS test score inferences was explored by examining AUC statistics from ROC analyses. The AUC demonstrates a ratio of sensitivity and specificity for identifying at-risk

Table 2
Correlation Coefficients Among Parent and Teacher IRS Scores and Parent and Teacher BASC-2 Scores (Study 1)

Variable	Parent-rated Externalizing Problems	Parent-rated Internalizing Problems	Parent-rated Adaptive Skills	Teacher-rated Externalizing Problems	Teacher-rated Internalizing Problems	Teacher-rated Adaptive Problems
Parent IRS (within and across raters)						
Peer relations	.45**	.28**	−.36**	.12**	.10*	−.12**
Parent–child relations	.46**	.24**	−.26**	.09*	.00	−.07
Academic performance	.41**	.23**	−.35**	.17**	.12**	−.13**
Self-esteem	.30**	.32**	−.26**	.04	.05	−.04
Family functioning	.44**	.26**	−.27**	.06	−.00	−.04
Overall	.49**	.19**	−.39**	.16**	.07	−.11**
Teacher IRS (within and across raters)						
Peer relations	.27**	.04	−.28**	.67*	.34*	−.49**
Teacher–child relations	.27**	.02	−.25**	.66*	.32*	−.48**
Academic performance	.24**	.03	−.30**	.50*	.25*	−.57**
Self-esteem	.22**	.04	−.24**	.48*	.46*	−.54**
Classroom functioning	.31**	.02	−.27**	.72*	.24*	−.49**
Overall	.25**	.03	−.29**	.60*	.31*	−.55**

Note. Nonparametric correlations were also computed using Spearman's rho, and a similar pattern of results was obtained. IRS = Impairment Rating Scale; BASC-2 = Behavioral Assessment System for Children, second edition.
* $p < .05$. ** $p < .01$.

status for each value rating on the IRS. AUCs were calculated for each IRS domain and ROC analyses were conducted separately by teacher-based BASC-2 risk status and parent-based BASC-2 risk status (see Table 3).

The AUCs for teacher IRS scores identifying teacher-based BASC-2 risk status ranged from .79 to .87, suggesting moderate to strong diagnostic validity of teacher IRS test score inferences. The AUCs for the parent IRS ratings identifying teacher-based BASC-2 risk status were significantly lower and not better than chance, ranging from .50 to .53. The AUCs for parent IRS scores identifying parent-based BASC-2 risk status ranged from .62 to .66, suggesting low to moderate diagnostic validity of parent IRS test score inferences. The AUCs for the teacher IRS scores identifying parent-based BASC-2 risk status were lower in magnitude but greater than chance, ranging from .58 to .60. The above analyses were also conducted with each of the BASC-2 composite scores; the pattern of AUCs was similar to the above with regard to rater (i.e., within-rater validity stronger than cross-rater) and magnitude within rater.

Statistical analyses (Hanley & McNeil, 1983) were also conducted to examine whether any one IRS domain *within rater* had a significantly higher AUC than others. For teacher IRS scores identifying teacher-based BASC-2 risk status, the AUCs for peer relationships, self-esteem, classroom functioning, and overall impairment were all significantly higher than the AUC for teacher–child relationships (all $ps < .05$; see superscripts in Table 3). No significant differences were detected for the parent IRS ratings identifying parent-based risk status.

Aim 3: Diagnostic efficiency. To examine what level of impairment on the parent and teacher IRS best differentiates typical and at-risk children, diagnostic efficiency statistics were examined for cutoff scores of 2, 3, and 4 on (a) the parent-rated IRS identifying parent-based BASC-2 risk status, (b) parent-rated IRS identifying teacher-based BASC-2 risk status, (c) teacher-rated IRS identifying teacher-based BASC-2 risk status, and (d) teacher-rated IRS identifying parent-based BASC-2 risk status (see Table 4). Because not one single domain on the parent or teacher IRS that had higher AUCs than *all* other domains, the overall impairment score on the parent and teacher IRS were used for these analyses. Further, the base rates for at-risk status in the current sample were used when calculating these statistics (28.2% for teacher-based BASC-2 risk status, 20.6% for parent-based BASC-2 risk status).

Table 3
Area Under the Curve (AUC) for Parent and Teacher IRS Scores Identifying Teacher-Based and Parent-Based At-Risk Status by BASC-2 Composites (Study 1)

	Teacher-based BASC-2 at-risk status (n = 117)			Parent-based BASC-2 at-risk status (n = 153)		
IRS domain	AUC	*SE*	95% CI	AUC	*SE*	95% CI
Parent IRS						
Peers	.51	.03	[.45, .57]	**.66**	**.03**	**[.60, .71]**
Parent–child	.51	.03	[.45, .57]	**.62**	**.03**	**[.57, .68]**
Academic	.53	.03	[.47, .60]	**.63**	**.03**	**[.57, .68]**
Self-esteem	.53	.03	[.46, .58]	**.63**	**.03**	**[.58 .69]**
Family	.50	.03	[.45, .56]	**.63**	**.03**	**[.58, .69]**
Overall	.53	.03	[.47, .59]	**.63**	**.03**	**[.58, .69]**
Teacher IRS						
Peers	**.85**[a]	**.02**	**[.81, .90]**	.58	.03	[.53, .64]
Teacher–child	**.79**[b]	**.03**	**[.74, .85]**	.58	.03	[.53, .63]
Academic	**.81**[a,b]	**.03**	**[.76, .86]**	.59	.03	[.53, .64]
Self-esteem	**.85**[a]	**.02**	**[.79, .89]**	.60	.03	[.54, .65]
Classroom	**.84**[a]	**.02**	**[.81, .89]**	.59	.03	[.54, .65]
Overall	**.87**[a]	**.02**	**[.83, .91]**	.59	.03	[.54, .64]

Note. AUCs within the same column within rater that have different superscripts are significantly different ($p < .05$) in identifying risk status. Children are considered at risk if they received a *T* score of 60 or higher on either the Externalizing Problems or Internalizing Problems Composites, or a score of 40 or lower on the Adaptive Skills Composite of the Behavioral Assessment System for Children, second edition, for that rater. IRS = Impairment Rating Scale; BASC-2 = Behavioral Assessment System for Children, second edition; CI = confidence interval. Boldface type denotes within-rater analyses.

Summary and Limitations

There is wide variability in the qualitative terms used to describe the magnitude of AUCs (Rice & Harris, 1995; Swets, 1996; Tape, n.d.). Taking into account this variability, our results indicate that there is moderate to strong concurrent and diagnostic validity for teacher IRS test score inferences when the criterion is teacher-rated BASC-2 scores. Parent IRS test score inferences showed low to moderate concurrent validity and diagnostic validity with parent-rated BASC-2 scores. Cross-rater validity was limited. Consistent with our hypothesis, the results for Aim 2 indicate that scores of 3 or 4 serve as a reasonable threshold for determining risk status as rated by the BASC-2. Following the presentation of Study 2 results, implications of each of these cut scores are discussed.

Study 2 was designed to address the limitations of Study 1. First, although the sample size in Study 1 is large (N = 568), the screening did not include all enrolled kindergarteners (63% consent rate), which reduces confidence that results are representative of the population. Second, although the BASC-2 scores were chosen as the criterion measure against which to compare IRS scores, this version of the BASC-2 was not designed for screening. Third, the original IRS instructions state, "Please mark an 'X' on the lines at the points that you believe reflect the impact of the child's problems on this area and whether he or she needs treatment or special services for the problems." Some parents had a negative reaction if they did not perceive their child as having problems or in need of services. Thus, the instructions were modified in Study 2 to better align with screening purposes. Lastly, we examined in Study 1 concurrent validity of IRS score inferences with other rating scales only. Research examining the validity of IRS test score inferences in relation to other kindergarten outcomes (i.e., grades, test scores, daily behavioral functioning, academic performance) was warranted. These limitations were addressed in Study 2.

Study 2

Method

Participants. Two years following completion of Study 1, one school district from Study 1 (with five elementary schools) participated in Study 2. The parents of 273 kindergarteners (94% of all enrolled) consented to participate, along with their teachers (n = 12; 100% female; 100% Caucasian). Of the 273, 242 students had complete parent IRS and teacher IRS and BESS rating scales

Table 4
Diagnostic Efficiency Statistics for Parent and Teacher Overall IRS Ratings Identifying Parent-Based and Teacher-Based BASC-2 Risk Status (Study 1)

Variable	SEN	SPE	EFF	PPV	NPV	TP (%)	TN (%)	FP (%)	FN (%)
Parent-rated overall impairment identifying parent-based BASC-2 risk status (base rate = 28.2%)									
Cutoff score of 2	.25	.94	.74	.61	.76	6.6%	67.5%	4.3%	21.5%
Cutoff score of 3	.14	.98	.74	.68	.74	3.8%	70.0%	1.8%	24.3%
Cutoff score of 4	.10	.99	.74	.95	.74	2.8%	71.7%	0.1%	25.4%
Teacher-rated overall impairment identifying parent-based BASC-2 risk status (base rate = 28.2%)									
Cutoff score of 2	.35	.79	.67	.40	.76	9.9%	56.7%	14.9%	18.3%
Cutoff score of 3	.28	.86	.69	.43	.75	7.7%	61.3%	10.4%	20.4%
Cutoff score of 4	.17	.91	.70	.43	.74	4.8%	65.3%	6.4%	23.2%
Parent-rated overall impairment identifying teacher-based BASC-2 risk status (base rate = 20.6%)									
Cutoff score of 2	.15	.90	.75	.28	.80	2.9%	71.6%	7.8%	17.6%
Cutoff score of 3	.09	.95	.77	.32	.80	1.8%	75.6%	3.8%	18.8%
Cutoff score of 4	.05	.98	.79	.37	.80	1.1%	77.6%	1.8%	19.5%
Teacher-rated overall impairment identifying teacher-based BASC-2 risk status (base rate = 20.6%)									
Cutoff score of 2	.76	.88	.86	.63	.93	15.7%	70.2%	9.2%	4.9%
Cutoff score of 3	.57	.92	.85	.65	.89	11.8%	73.1%	6.4%	8.8%
Cutoff score of 4	.42	.97	.86	.78	.87	8.6%	76.9%	2.5%	12.0%

Note. Data represent children as either parent-based or teacher-based risk status on Externalizing if they received a *T* score of 60 or higher on the Externalizing Problems BASC-2 Composite. IRS = Impairment Rating Scale; BASC-2 = Behavioral Assessment System for Children, second edition; SEN = sensitivity; SPE = specificity; EFF = efficiency; PPV = positive predictive value; NPV = negative predictive value; TP = true positives; TN = true negatives; FP = false positives; FN = false negatives.

and were included in the analyses. Those not included in the analyses did not differ from those excluded with regard to age, gender, race, school building assignment, and mother's or father's highest level of education. Child participants were 50.8% male and 95.5% Caucasian (2.1% classified as Hispanic; less than 1% African American, Asian, and American Indian/Alaskan Native) with a mean age of 5.61 (*SD* = .46). The following data represent the highest education achieved by mothers and fathers in the sample: less than high school completion: 7.0% for mothers, 7.4% for fathers; high school completion: 23.6% for mothers, 38.5% for fathers; associate's degree or some college: 45.9% for mothers, 28.1% for fathers; bachelor's degree: 8.7% for mothers, 5.4% for fathers; graduate degree: 4.1% for mothers, 1.2% for fathers). This is consistent with parent educational data from the region.

Measures.

Parent demographic questionnaire. Parents completed a demographic questionnaire similar to that completed by parents in Study 1. It included information about child gender, age, race/ethnicity, and maternal and paternal education.

BASC-2 BESS. The BESS is a brief measure (26 items for teachers) designed to screen for behavioral and emotional problems in children (Kamphaus & Reynolds, 2007). The teacher version of the child/adolescent form (K–12th) was used as the criterion measure in Study 2. This form was normed on a large sample closely matched to the U.S. population with regard to race/ethnicity; scores have strong test–retest (>.85), interrater (≥.70), and internal reliability (≥.94), as well as evidence of convergent, divergent, and predictive validity for standardized reading and math scores, and student GPAs over 4 years (significant correlations range from .34 to .63; Kamphaus & Reynolds, 2007).

The IRS. Parents and teachers completed the IRS (description in Study 1). Again, the six parallel items on the parent and teacher versions were examined. Instructions were modified to acknowledge that parents may not perceive their child to have a problem or need services. The new instructions stated,

> The following questionnaire is intended to assess potential problems that your child may or may not currently experience. If you believe that your child is not experiencing social, emotional, or behavioral problems, make an "X" at the end of the line marked "no problem." If you believe that your child is experiencing any social, emotional, or behavioral problems, make an "X" on the line at the point that you believe reflects the impact of the child's problems in this area and whether he or she needs treatment or special services.

Academic outcomes. A quarterly grade-point average (GPA) was calculated using a 4-point scale based on participants' grades in Reading, Spelling, Math, Science, and Social Studies. Grades were coded as follows: above satisfactory (S+ = 4.0), satisfactory (S = 3.0), below satisfactory (S− = 2.0), need improvement (N = 1.0), and unsatisfactory (U = 0.0). Because results from Quarters 2 and 3 were similar to those in Quarter 4, for efficiency, only data for Quarters 1 and 4 are presented. Because all other criterion measures are, to some degree, dependent on teacher ratings (and possibly bias), students' scores on the school district's test of state-based kindergarten content standards for literacy/reading in Quarters 1 and 4 were also obtained. Although these are not state- or nationally normed tests, and they do not have known psychometric properties, they represent an ecologically valid indicator of student progress in kindergarten competencies, as these tests scores are used to make academic programming decisions within the district. Depending on the quarter, these tests include 18–26 items, each of which are scored as correct or incorrect. Scores are summed and the percent correct is calculated. According to school district professionals, a score below 80% correct is considered at risk. In sum, GPA and reading test scores for Quarters 1 and 4 represent the academic criterion measures in this Study 2.

Behavioral outcomes. All participating schools use a school-wide positive behavior support framework for discipline. Each teacher uses a behavior wheel that documents the student's rule-following behavior, and each student has an individual clothes pin attached to the wheel. Each student begins the day on the green segment of the wheel and moves his or her clothes pin with each additional rule violation, such that the yellow segment represents a first warning, the orange segment represents a second warning (with a possible mild consequence), and the red segment represents a referral to the office and parent notification. At the end of each day, the student documents on a monthly calendar the color that he or she achieved for that day. The percent of green days achieved for each student was calculated by dividing the total number of days the student attended school by the total number of green days achieved by the child. A child was considered at risk for behavioral problems if he or she fell below the threshold of achieving 80% green days. This decision was made in collaboration with school staff. They observed that it is not uncommon for typically developing students to earn one day off green (80% green; 4 out of 5 days). However, if a student falls below this 80% threshold (2 or more days off green), it raises concern about the child's difficulties with behavioral control and other associated problems. This score represents the behavioral criterion measure in this Study 2.

Procedure. This study was approved by the IRB. Parents were recruited to participate when registering their child for kindergarten (between April and August). Following consent, parents completed the demographic questionnaire and the parent IRS. Parents received a small educational gift (e.g., flashcards, alphabet magnets) as compensation for participation. Teachers consented to participate and completed the teacher IRS and BESS in October, and were compensated with $80.

Data preparation.

Definition of at risk. For the teacher BESS, age-based *T* scores were used to determine the classification of risk status according to the BESS manual and scoring program; namely, a *T* score of 61 or higher indicates "at risk" (teacher-based BESS risk status). In Study 2, 26 (10.7%) of the children were categorized as at risk according to the teacher BESS. For GPA, a child was considered "at risk" for academic problems if his or her GPA fell below 3.0; across quarters, 8%–18% of students fell below this cutoff. For reading tests scores, at-risk status was defined by a percent correct score lower than 80%; across quarters, 14.3%–29.1% of students fell below this cutoff. Lastly, for daily behavior data, a child was considered at risk for behavioral problems if he or she fell below the threshold of achieving 80% green days; 12% of the sample fell below this cutoff score.

The IRS. Level of risk on each domain of the IRS was assessed separately for parent and teacher ratings. Consistent with Study 1, diagnostic efficiency statistics for cutoff scores of 2, 3, and 4 were examined.

Results

Descriptive statistics for the parent- and teacher-rated IRS domains for the total sample and for teacher-based BESS at-risk and typical samples are presented in Table 5.

Aim 1: Concurrent validity. To assess concurrent validity, bivariate correlations between the scores on the parent and teacher IRS and the teacher BESS forms were examined (see Table 6). There was a lack of association or low correlations for parent-rated IRS scores and teacher-rated BESS scores (*r*s all lower than .24). The teacher-rated IRS scores yielded moderate to strong associations with the teacher-rated BESS score (*r*s range from .59 to .71). As expected, relations were stronger for within-rater scores (teacher IRS and teacher BESS) than between-rater scores (parent IRS and teacher BESS). The parent-rated IRS scores showed a lack of association or low associations with grades, reading tests scores, and behavior problems (*r*s all lower than .26). The teacher-rated IRS scores yielded moderate associations with first-quarter grades (*r*s range from −.29 to −.56), fourth-quarter grades (*r*s range

Table 5
IRS Scores for Kindergartners by Parent- and Teacher-Based Risk Status According to the BESS (Study 2)

	Total sample (N = 242)				Teacher-based at risk (n = 26)		Teacher-based typical (n = 216)		
IRS domain	*M*	*SD*	Skewness (*SE*)	Kurtosis (*SE*)	*M*	*SD*	*M*	*SD*	Effect Size
Parent IRS scores									
Peers relationship	0.42	1.02	3.21 (.16)	11.17 (.31)	0.51	1.16	0.40	0.99	.10
Parent–child relationship	0.44	1.06	3.00 (.16)	9.05 (.31)	0.52	1.23	0.42	1.02	.09
Academic performance	0.37	0.90	3.22 (.16)	11.61 (.32)	0.50	1.02	0.33	0.87	.18
Self-esteem	0.41	0.99	3.40 (.16)	13.50 (.31)	0.44	0.98	0.41	0.99	.03
Family functioning	0.43	1.04	3.20 (.16)	11.32 (.31)	0.41	1.00	0.43	1.06	−.02
Overall impairment	0.44	1.11	2.68 (.16)	7.23 (.31)	0.58	1.32	0.41	1.05	.14
Teacher IRS scores									
Peers relationship	0.83	1.40	2.39 (.16)	5.98 (.31)	2.52*	1.77	0.39*	0.86	1.53
Teacher–child relationship	0.71	1.37	2.27 (.16)	5.75 (.31)	2.19*	1.94	0.33*	0.82	1.25
Academic performance	1.08	1.73	1.92 (.16)	3.29 (.31)	2.91*	2.08	0.61*	1.26	1.34
Self-esteem	0.93	1.49	2.19 (.16)	4.73 (.31)	2.59*	1.77	0.49*	1.04	1.45
Classroom functioning	0.85	1.52	2.14 (.16)	4.89 (.31)	2.67*	1.99	0.38*	0.91	1.48
Overall impairment	1.00	1.63	1.88 (.16)	3.20 (.31)	2.99*	1.86	0.48*	1.08	1.65

Note. Effect sizes were calculated using Cohen's *d* (Cohen, 1988). IRS = Impairment Rating Scale; BESS = Behavioral and Emotional Screening System.
* $p < .01$ (significant differences between the at-risk and typical groups).

Table 6
Correlation Coefficients Among Parent and Teacher IRS Scores and Teacher-Based BESS, Quarterly Grades, Tests Scores, and Daily Behavior (Study 2)

Variable	Teacher-based BESS	First-quarter grades	Fourth-quarter grades	First-quarter reading tests	Fourth-quarter reading tests	Behavior problems
Parent IRS						
Peer relations	.23**	−.14*	−.15*	.10	.04	.25**
Parent–child relations	.14*	−.07	−.11	.08	.07	.16*
Academic performance	.21*	−.08	−.03	.10	−.02	.23**
Self-esteem	.07	.00	−.00	.05	.01	.03
Family functioning	.12	−.03	−.09	.04	.02	.15*
Overall functioning	.22*	−.14*	−.13*	.12	.06	.20**
Teacher IRS						
Peer relations	.59**	−.30**	−.25**	.15*	.24***	.63**
Teacher–child relations	.64**	−.29**	−.26**	.11	.19**	.58**
Academic performance	.69**	−.56**	−.47**	.38***	.40**	.60**
Self-esteem	.67**	−.33**	−.30**	.15*	.22**	.68**
Classroom functioning	.62**	−.43**	−.38**	.23***	.35***	.55**
Overall functioning	.71**	−.47**	−.46**	.29***	.37***	.65**

Note. Nonparametric correlations were also computed using Spearman's rho, and a similar pattern of results was obtained. IRS = Impairment Rating Scale; BESS = Behavioral and Emotional Screening System.
* $p < .05$. ** $p < .01$. *** $p < .001$.

from −.25 to −.47), first-quarter reading test scores (*r*s range from .11 to .38), fourth-quarter reading test scores (*r*s range from .19 to .40), and behavior problems (*r*s range from .55 to .68).

Aim 2: Discriminant validity. To measure the extent to which scores on the IRS discriminated between at-risk and typical groups as determined by the BESS, the diagnostic validity of the parent and teacher IRS test score inferences was explored by examining AUC statistics from ROC analyses. AUCs were calculated for each IRS domain (see Table 7). The AUCs for teacher IRS scores identifying teacher-based BESS risk status ranged from .86 to .93, suggesting moderate to strong diagnostic validity of the teacher IRS test score inferences. The AUCs for the parent IRS test score inferences identifying teacher-based BESS risk status suggested low to moderate diagnostic validity (ranging from .55 to .66). Statistical analyses (Hanley & McNeil, 1983) were conducted to examine whether any one IRS domain *within rater* had a significantly higher AUC than others. No significant differences were detected for the teacher IRS ratings identifying teacher-based BESS risk status.

The AUCs for the teacher IRS scores identifying first-quarter grades risk status and test score risk status (.57–.68) and fourth-quarter grades and test score risk status (.65–.75) suggested low to

Table 7
Area Under the Curve (AUC) for Parent and Teacher IRS Scores Identifying Teacher-Based At-Risk Status by BESS Composite, Grades, Test Scores, and Behavior (Study 2)

	Teacher-based BESS at-risk status (n = 26)			First-quarter grades at-risk status (n = 37)			Fourth-quarter grades at-risk status (n = 24)			First-quarter reading scores (n = 71)			Fourth-quarter reading scores (n = 45)			Behavior at-risk status (n = 27)		
IRS domain	AUC	*SE*	95% CI	AUC	*SE*	95% CI	AUC	*SE*	95% CI	AUC	*SE*	95% CI	AUC	*SE*	95% CI	AUC	*SE*	95% CI
Parent IRS																		
Peers	.62[a,b]	.07	[.49, .75]	.52	.05	[.41, .62]	.53	.06	[.40, .65]	.54	.04	[.46, .62]	.54	.05	[.44, .64]	.58	.06	[.45, .71]
Parent–child	.58[a,b]	.07	[.45, .71]	.53	.05	[.42, .63]	.56	.07	[.43, .68]	.53	.04	[.45, .61]	.56	.05	[.46, .65]	.56	.06	[.43, .68]
Academic	.66[b]	.07	[.53, .78]	.54	.05	[.43, .64]	.51	.06	[.39, .63]	.55	.04	[.46, .63]	.50	.05	[.40. .60]	.60	.07	[.47, .72]
Self-esteem	.55[a]	.07	[.42, .68]	.48	.05	[.38, .58]	.49	.06	[.37, .62]	.50	.04	[.42, .59]	.52	.05	[.42, .62]	.49	.06	[.37, .61]
Family	.62[a,b]	.07	[.49, .75]	.50	.05	[.39, .60]	.53	.06	[.40, .66]	.51	.04	[.43, .59]	.56	.05	[.46, .65]	.54	.06	[.42, .66]
Overall	.66[b]	.07	[.53, .78]	.56	.05	[.45, .67]	.54	.07	[.41, .67]	.57	.04	[.48, .65]	.57	.04	[.48, .66]	.57	.06	[.44, .69]
Teacher IRS																		
Peers	.86	.05	[.77, .95]	.64[a]	.05	[.55, .74]	.65[a]	.06	[.54, .77]	.57[a]	.04	[.49, .66]	.69[a,b]	.05	[.59, .78]	.83[a]	.05	[.73, .92]
Teacher–child	.86	.05	[.77, .96]	.64[a]	.05	[.54, .74]	.66[a]	.06	[.54, .77]	.57[a]	.04	[.48, .66]	.65[a]	.05	[.56, .74]	.82[a]	.05	[.73, .91]
Academic	.92	.02	[.87, .97]	.77[b]	.05	[.68, .87]	.79[b]	.05	[.70, .89]	.68[b]	.04	[.60, .76]	.75[b]	.04	[.66, .84]	.85[a,b]	.05	[.75, .94]
Self-esteem	.92	.03	[.87, .97]	.65[a]	.05	[.56, .75]	.66[a]	.06	[.55, .78]	.59[a]	.04	[.51, .67]	.67[a]	.05	[.58, .76]	.87[a,b]	.04	[.80, .95]
Classroom	.89	.04	[.81, .97]	.70[a]	.05	[.60, .80]	.69[a]	.06	[.57, .82]	.58[a]	.04	[.50, .67]	.71[a,b]	.05	[.62, .80]	.83[a]	.05	[.74, .92]
Overall	.93	.02	[.89, .98]	.75[a,b]	.04	[.67, .84]	.79[b]	.05	[.69, .88]	.66[a,b]	.04	[.58, .74]	.74[b]	.04	[.65, .82]	.89[b]	.03	[.83, .95]

Note. AUCs within the same column within rater that have different superscripts are significantly different ($p < .05$) in identifying risk status. Children are considered at risk if they received a *T* score of 60 or higher on either the Externalizing Problems or Internalizing Problems Composites, or a score of 40 or lower on the Adaptive Skills Composite of the BASC-2 for that rater. IRS = Impairment Rating Scale; BESS = Behavioral and Emotional Screening System; CI = confidence interval.

moderate diagnostic validity of the teacher IRS test score inferences. The AUCs for the teacher IRS ratings identifying behavior risk status ranged from .82 to .89, suggesting moderate to strong diagnostic validity of the teacher IRS test score inferences in identifying classroom behavior problems. The AUCs for the parent IRS ratings identifying risk status based on first- and fourth-quarter grades and reading tests scores or daily behavior reports were significantly lower and not much better than chance, ranging from .48 to .60.

Aim 3: Diagnostic efficiency. To examine which level of impairment on the parent and teacher IRS best differentiates typical and at-risk children according to the BESS, diagnostic efficiency statistics were examined for cutoff scores of 2, 3, and 4 on (a) teacher-rated IRS identifying teacher-based BESS risk status and (b) the parent-rated IRS identifying teacher-based BESS risk status (see Table 8). Consistent with Study 1, there was no single domain on the parent or teacher IRS that had higher AUCs than *all* other domains. Thus, the overall impairment score on the parent and teacher IRS were used for these analyses. Further, the base rates for at-risk status on the BESS in the current sample were used when calculating these statistics (10.7%; see Table 8).

Discussion

Taken together, the results of Studies 1 and 2 suggest that the test score inferences of the teacher IRS have moderate to strong validity and the test scores inferences of the parent IRS have low to moderate validity in determining risk for social, emotional, academic, and behavioral problems. These preliminary data suggest that the teacher IRS has utility as a screening tool. Evidence for utility of the IRS was found with teacher ratings on the BASC-2 and BESS, as well as with other important kindergarten outcomes (i.e., GPA, reading test scores, and daily behavior). As anticipated, results from both studies showed stronger concurrent validity within rater, than across rater. Further, the findings from both studies suggest that a cutoff score of 3 or 4 may serve as a reasonable threshold for determining risk status as rated by the BASC-2 and BESS, depending on the goal of the screening (e.g., to monitor or to further assess). Lastly, the second study demonstrated evidence of ecological validity; IRS scores were moderately related to important academic indicators (GPA and reading test scores), as well as highly predictive of daily outcomes from a classwide behavioral discipline program. These studies offer preliminary evidence that the brief, publically available teacher IRS may be useful as a universal screening tool for school districts concerned about time and cost. Given the state of multiply-stressed school systems and the dismal rate (2%) at which screenings occur in schools across the country (Romer & McIntosh, 2005), these results are encouraging. The implications for research and practice are discussed below.

The outcomes of both studies provided encouraging evidence for the teacher IRS. These data are consistent with other studies documenting the utility of teacher ratings in predicting academic and behavior outcomes (DiStefano & Kamphaus, 2007; Kamphaus et al., 2007; Kamphaus et al., 2010). However, caution may be warranted for the parent IRS. On one hand, the low to moderate validity of test score inferences among parent ratings is not surprising given the research showing low correspondence between parent and teacher ratings of child behavior (Achenbach et al., 1987). It is possible that weaker validity of test score inferences among parent ratings and differences in agreement across informants could be attributed to contextual factors. Namely, the settings in which teachers observe children may be significantly more structured and more academically and socially demanding than those in which parents observe young children. These different contexts may produce different child behaviors. Thus, parents and teachers may be reporting on behaviors that only occur in one of the settings. Alternatively, these differences may be a function of sampling at different time points. In fact, in both studies, parents rated children prior to the start of kindergarten, whereas teachers rated children 8–12 weeks into the kindergarten year. However, the weaker cross-informant predictions may be a function of the screening tool itself, as there is emerging data that parent ratings on other screening measures have incremental validity in predicting important kindergarten outcomes, beyond that explained by the academic screener used by the school district (Owens et al., 2014). Thus, the cross-informant data from these studies should not necessarily be evidence for dismissing the utility of parent ratings at kindergarten entry. Additional research on the validity on parent IRS test score inferences is indicated, with different samples and outcomes, especially as the use of multiple informants is recommended for evidence-based assessment, and can offer unique information about setting-specific behavior (Pelham, Fabiano, & Massetti, 2005). Further, even if parent IRS scores are only moderately predictive of teacher ratings or other objective kindergarten

Table 8
Diagnostic Efficiency for Parent and Teacher Overall IRS Ratings Identifying Parent-Based and Teacher-Based BESS Risk Status (Study 2)

Variable	SEN	SPE	EFF	PPV	NPV	TP (%)	TN (%)	FP (%)	FN (%)
Parent-rated overall impairment identifying teacher-based BESS teacher risk status (base rate = 10.74%)									
Cutoff score of 2	.29	.91	.85	.29	.92	3.2%	74.8%	7.7%	7.0%
Cutoff score of 3	.17	.95	.86	.28	.90	1.8%	77.7%	4.5%	9.0%
Cutoff score of 4	.08	.97	.87	.22	.90	0.9%	78.9%	3.2%	9.9%
Teacher-rated overall impairment identifying teacher-based BESS risk status (base rate = 10.74%)									
Cutoff score of 2	.80	.90	.89	.50	.97	8.3%	80.8%	8.8%	2.1%
Cutoff score of 3	.72	.95	.93	.65	.97	7.5%	85.4%	4.2%	2.9%
Cutoff score of 4	.52	.98	.93	.77	.94	5.4%	87.9%	1.7%	5.0%

Note. IRS = Impairment Rating Scale; BESS = Behavioral and Emotional Screening System; SEN = sensitivity; SPE = specificity; EFF = efficiency; PPV = positive predictive value; NPV = negative predictive value; TP = true positives; TN = true negatives; FP = false positives; FN = false negatives.

outcomes, there may be utility in the *process* of having parents and teacher both complete ratings at kindergarten entry. Namely, parent receipt of a report providing feedback of the screening may facilitate early communication between parents and teachers (Girio-Herrera, Owens, & Langberg, 2014) and monitoring of the problematic behavior, possibly fostering early service engagement among parents.

Thus, given the promising results obtained in this study, further examination of the role of the IRS in a multistep screening process may be fruitful. For example, school professionals could use the parent IRS to obtain a preliminary profile of children at kindergarten entry. Consistent with previous research (Fabiano et al., 2006), scores of 3 or 4 on the IRS seem to be appropriate cutpoints for detecting a broad definition of risk status. Such data may help principals distribute higher risk children across general education classrooms, as well as identify children who warrant early monitoring. Because teacher test score inferences have greater diagnostic utility than parent test score inferences, a parent score of 3 or higher could simply be used to trigger monitoring. However, teacher scores obtained at the end of the first grading period may trigger additional activities, such as a specific parent–teacher team meeting, referral to early intervention program, and/or additional assessment. Ultimately, the role of the screener and an optimal cutpoint should be determined on the basis of the school district's goals, intended purpose of screening, and the availability of other tools and resources. Further, school professionals must consider the rates of false negatives and false positives resulting from IRS scores. Longitudinal data are needed to fully understand the characteristics of children who are and are not identified by the IRS. Until then, for example, if the IRS is to be used as a universal screening tool only at kindergarten entry, school professionals may be more willing to accept a higher rate of false negatives (children at risk who go undetected) than a higher rate of false positives (typical children who are incorrectly identified as at risk) because the "missed" children will likely ultimately be detected at a later point in time and because there may be negative consequences of providing inaccurate information implying risk status to parent of a child during his or her first year of schooling.

Study 2 included an examination of the IRS scores in predicting "real world" academic and behavioral indicators. Teacher IRS test score inferences had moderate diagnostic validity with first- and fourth-quarter grades and reading test scores and with daily behavior outcomes on a classwide discipline system. These results are valuable to our scientific community in evaluating this promising screener; however, to be able to predict fourth-quarter GPA (highest AUC = .79) and daily behavior throughout the year (highest AUC = .89) via a six-item rating scale is of great interest to our school partners. Namely, school administrators and teachers may find this information useful for identifying students who may benefit from monthly review on school-based support teams (e.g., positive behavior support teams, intervention assistance teams), identifying students in need of preventative academic or behavioral support, preparing for earlier discussions with parents, and/or matching specific students with particular teachers. It is possible that providing the schools with this information may motivate school systems to participate in mental health screening.

Although these data are promising for identifying a brief, free, publically available tool, additional research on the infrastructure needed for schools to adopt the IRS as a screener is warranted. For example, future studies may examine (a) processes that help schools take ownership of the screening process; (b) the types of trainings (e.g., one time vs. periodic) that produce maximum understanding and use of the data by administrators, principals, and teachers; and (c) methods for communicating with parents about the results. Obtaining feedback from school administrators and staff, as well as parents, regarding the feasibility, utility, and acceptability of the IRS has critical implications for its adoption and sustainability. Given the preliminary nature of the current data, it is recommend that schools not proceed with the IRS unless doing so in partnership with a research team, as additional examination of this instrument is needed.

These studies were not without limitations. First, the samples were limited to those schools with administrators who agreed to participate, thus making it a sample of convenience. Given the response rates and the sample characteristics (almost entirely Caucasian and living in low-income, rural communities), the results may not generalize to the entire population from which the sample was drawn or to other samples with differing characteristics. As such, we encourage caution in adopting the IRS in practice until further study using racially and economically diverse samples occurs. Similarly, given that both studies were limited to one geographic location, future research examining the IRS should include a nationally representative sample. Second, the validity of the inferences derived from the IRS in these studies may differ from other studies if other criterion measures are used (other than the BASC-2 and BESS) or populations examined. Third, the use of 80% as the cutpoint for our behavioral outcome in Study 2 was determined in collaboration with our school partners in this particular school district. Alternative thresholds may be more meaningful in different school districts and should be explored in future research. Lastly, although there are advantages to single-item measures, particularly in screening and epidemiological research (i.e., reduced burden and cost, and ease of interpretation; see Bowling, 2005, for discussion), there are also weaknesses. Namely, the single-item nature of the IRS may make it less stable, reliable, and precise (i.e., may not fully capture the construct of interest) than multiitem scales. Thus, replication is warranted to enhance confidence in the IRS as a screening tool.

Conclusion

In summary, the psychometric and ecological utility of both parent and teacher versions of the IRS are worthy of continued research. Correlational analyses, as well as ROC and diagnostic efficiency analyses, indicate that IRS scores, particularly teacher IRS scores, have moderate to strong psychometric properties associated with identifying children who are at risk for a broad range of social, emotional, academic, or behavioral problems. The brevity and public availability of the IRS make it a time- and cost-efficient measure that school districts may find appealing. Given the high rates of undetected mental health problems in our county, as well as the benefits of universal screening and early intervention, identification of a tool such as the IRS has important utility in both research and practice.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-

informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232. doi:10.1037/0033-2909.101.2.213

Albers, C. A., Glover, T. A., & Kratochwill, T. R. (2007). Introduction to the special issue: How can universal screening enhance educational and mental health outcomes? *Journal of School Psychology, 45,* 113–116. doi:10.1016/j.jsp.2006.12.002

Ashford, J., Smit, F., van Lier, P. A., Cuipers, P., & Koot, H. M. (2008). Early risk indicators of internalizing problems in late childhood: A 9-year longitudinal study. *Journal of Child Psychology and Psychiatry, 49,* 774–780. doi:10.1111/j.1469-7610.2008.01889.x

Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health, 59,* 342–345.

Brady, C. E., Evans, S. W., Berlin, K. S., Bunford, N., & Kern, L. (2012). Evaluating school impairment with adolescents using the Classroom Performance Survey. *School Psychology Review, 41,* 429–446.

Briggs-Gowan, M. J., Horwitz, S. M., Schwab-Stone, M. E., Leventhal, J M., & Leaf, P. (2000). Mental health in pediatric settings: Distribution of disorders and factors related to service use. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 841–849.

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

DiStefano, C., & Kamphaus, R. W. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders, 15,* 93–102. doi:10.1177/10634266070150020401

Dowdy, E., & Kamphaus, R. W. (2007). A comparison of classification methods for use in predicting school-based outcomes. *The California School Psychologist, 12,* 121–132. doi:10.1007/BF03340936

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. *School Mental Health, 2,* 166–176. doi:10.1007/s12310-010-9036-3

Drummond, T. (1993). *The Student Risk Screening Scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.

Eklund, K., Renshaw, T., Dowdy, E., Jimerson, S., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *California School Psychologist, 14,* 89–95. doi:10.1007/BF03340954

Evans, S. W., Brady, C. E., Harrison, J. R., Bunford, N., Kern, L., State, T., & Andrews, C. (2013). Measuring ADHD and ODD symptoms and impairment using high school teachers' ratings. *Journal of Clinical Child & Adolescent Psychology, 42,* 197–207. doi:10.1080/15374416.2012.738456

Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-Revised*. Odessa, FL: Psychological Assessment Resources.

Fabiano, G. A., Pelham, W. E., Waschbusch, D. A., Gnagy, E. M., Lahey, B. B., Chronis, A. M., . . . Burrows-Maclean, L. (2006). A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *Journal of Clinical Child & Adolescent Psychology, 35,* 369–385. doi:10.1207/s15374424jccp3503_3

Fabiano, G. A., Vujnovic, R. K., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Pariseau, M. E., . . . Volker, M. (2010). Enhancing the effectiveness of special education programming for children with attention deficit hyperactivity disorder using a daily report card. *School Psychology Review, 39,* 219–239.

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

Girio-Herrera, E. L., Owens, J. S., & Langberg, J. M. (2013). Perceived barriers to help-seeking among parents of at-risk kindergarteners in rural communities. *Journal of Clinical Child & Adolescent Psychology, 42,* 68–77. doi:10.1080/15374416.2012.715365

Girio-Herrera, E. L., Owens, J. S., & Langberg, J. M. (2014). *A pilot study examining a school-based parent engagement intervention following school mental health screening*. Manuscript submitted for publication.

Glascoe, F. (1999). The validation and standardization of parents' evaluation of developmental status. *Diagnostique, 23,* 185–203.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117–135. doi:10.1016/j.jsp.2006.05.005

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Halfon, N., Regalado, M., Sareen, H., Inkelas, M., Peck Reuland, C. H., Glascoe, F. P., & Olson, L. M. (2004). Assessing development in the pediatric office. *Pediatrics, 113,* 1926–1933.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843. doi:10.1148/radiology.148.3.6878708

Hill, L. G., Lochman, J. E., Coie, J. D., Greenberg, M. T., & The Conduct Problems Prevention Research Group. (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology, 72,* 809–820. doi:10.1037/0022-006X.72.5.809

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.

Kamphaus, R. W., & Reynolds, C. R. (2007). *Behavior Assessment System for Children–Second Edition (BASC-2): Behavioral and Emotional Screening System (BESS)*. Bloomington, MN: Pearson.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67,* 342–356. doi:10.1177/0013164407067002100l

King, T. M., & Glascoe, F. P. (2003). Developmental surveillance of infants and young children in pediatric primary care. *Current Opinion in Pediatrics, 15,* 624–629. doi:10.1097/00008480-200312000-00014

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45,* 163–191. doi:10.1016/j.jsp.2006.11.005

Little, M., Murphy, J., Jellinek, M., Bishop, S., & Arnett, H. (1994). Screening 4- and 5-year-old children for psychosocial dysfunction: A preliminary study with the Pediatric Symptom Checklist. *Journal of Developmental and Behavioral Pediatrics, 15,* 191–197.

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, 2,* 115–126. doi:10.1080/09362839109524774

Lochman, J. E., & The Conduct Problems Prevention Research Group. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology, 63,* 549–559. doi:10.1037/0022-006X.63.4.549

McGuire, J., & Richman, N. (1986). Screening for behavior problems in nurseries: The reliability and validity of the Preschool Behaviour Checklist. *Journal of Child Psychology and Psychiatry, 27,* 7–32. doi:10.1111/j.1469-7610.1986.tb00618.x

Owens, J. S., Murphy, C. E., Richerson, L., Girio, E. L., & Himawan, L. K. (2008). Science to practice in underserved communities: The effectiveness of school mental health programming. *Journal of Clinical Child & Adolescent Psychology, 37,* 434–447. doi:10.1080/15374410801955912

Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V., Watabe, Y., Himawan, L., . . . Andrews, N. (2014). *Screening for social, emotional,*

and behavioral problems at kindergarten entry: Diagnostic utility and predictive validity of parent report. Manuscript under review.

Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 34*, 449–476. doi:10.1207/s15374424jccp3403_5

Pihlakoski, L., Sourander, A., Aromaa, M., Rautava, P., Helenius, H., & Sillanpaa, M. (2006). The continuity of psychopathology from early childhood to preadolescence. *European Child & Adolescent Psychiatry, 15*, 409–417. doi:10.1007/s00787-006-0548-1

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children (BASC)*. Circle Pines, MN: American Guidance Service.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737–748. doi:10.1037/0022-006X.63.5.737

Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry, 155*, 715–725.

Romer, D., & McIntosh, M. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D. L. Evans, E. B. Foa, R. E. Gur, H. Hendin, C. P. O'Brien, M. Seligman, & B. T. Walsh (Eds.), *Treating and preventing adolescent mental health disorders* (pp. 598–615). Oxford, England: Oxford University Press.

Shaffer, D., Scott, M., Wilcox, H., Maslow, C., Hicks, R., Lucas, C., . . . Greenwald, S. (2004). The Columbia Suicide Screen: Validity and reliability of a screen for youth suicide and depression. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 71–79.

Squires, J., Bricker, D., & Twombly, E. (2002). *Ages and Stages Questionnaires: Social and emotional*. Baltimore, MD: Brooks Publishing.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnosis: Collected papers*. Mahwah, NJ: Erlbaum.

Tape, T. G. (n.d.). *Interpreting diagnostic tests*. Retrieved from http://gim.unmc.edu/dxtests/Default.htm

Tremblay, R. E., Masse, B., Perron, D., Leblanc, M., Schwartzman, A. E., & Ledingham, J. E. (1992). Early disruptive behavior, poor school achievement, delinquent behavior, and delinquent personality: Longitudinal analysis. *Journal of Consulting and Clinical Psychology, 60*, 64–72. doi:10.1037/0022-006X.60.1.64

U.S. Census Bureau. (2000). *Census Summary File 3, Table PCT23*. Retrieved from http://www.pewtrusts.org/news_room_detail.aspx?id=19570

U.S. Census Bureau. (2006). *State and county quick facts*. Retrieved from http://quickfacts.census.gov/qfd/states/39000.html

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13*, 79–96. doi:10.1177/10634266050130020201

Walker, H. M., & Severson, H. (1992). *Systematic Screening for Behavior Disorders: Technical manual*. Longmont, CO: Sopris West.

Received March 20, 2013
Revision received May 1, 2014
Accepted June 25, 2014 ■