School Psychology Quarterly
2014, Vol. 29, No. 4, 438–451

1045-3830/14/$12.00 http://dx.doi.org/10.1037/spq0000100

Development of a Problem-Focused Behavioral Screener Linked to Evidence-Based Intervention

Brian Daniels, Robert J. Volpe, and Amy M. Briesch
Northeastern University

Gregory A. Fabiano
University at Buffalo, State University of New York

This study examines the factor structure, reliability and validity of a novel school-based screening instrument for academic and disruptive behavior problems commonly experienced by children and adolescents with attention deficit hyperactivity disorder (ADHD). Participants included 39 classroom teachers from two public school districts in the northeastern United States. Teacher ratings were obtained for 390 students in grades K-6. Exploratory factor analysis supports a two-factor structure (oppositional/disruptive and academic productivity/disorganization). Data from the screening instrument demonstrate favorable internal consistency, temporal stability and convergent validity. The novel measure should facilitate classroom intervention for problem behaviors associated with ADHD by identifying at-risk students and determining specific targets for daily behavior report card interventions.

Keywords: attention deficit hyperactivity disorder (ADHD), behavioral assessment, daily behavior report card, classroom intervention, universal screening

Attention deficit hyperactivity disorder (ADHD) affects approximately 7–9% of children and adolescents in the United States (Visser, Lesesne, & Perou, 2007; Visser et al., 2014). Children with ADHD are at risk for several negative school-related outcomes including higher levels of disruptive and off-task behaviors, more interpersonal problems, lower average academic performance, and higher risk for special education services, grade retention, and dropout than typically developing students (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2007; McConaughy, Volpe, Antshel, Gordon, & Eiraldi, 2011; Wehmeier, Schacht, & Barkley, 2010). The potential for these outcomes indicates a significant need for accurate assessment and timely intervention to reduce the impact of problems associated with ADHD.

Best practices in ADHD assessment involve the use of multiple methods, although the specific methods employed depend on whether the purpose of assessment is screening, diagnosis, intervention planning or outcome evaluation (Pelham, Fabiano, & Massetti, 2005; Tobin, Schneider, Reck, & Landau, 2008). A common first step in screening school-age children for ADHD is to collect teacher ratings of student classroom behavior (DuPaul, 2004; DuPaul & Kern, 2011). Teacher ratings are efficient and feasible when compared with other methods and are considered an effective method to identify children at risk for behavioral difficulties (Dwyer, Nicholson, & Battistutta, 2006; Jones, Dodge, Foster, Nix, & the Conduct Problems Prevention Research Group, 2002).

Glover and Albers (2007) identified three important considerations for evaluating universal screening measures: (a) appropriateness, (b) technical adequacy, and (c) usability. For a

Brian Daniels, Robert J. Volpe, and Amy M. Briesch, Department of Counseling and Applied Educational Psychology, Northeastern University; Gregory A. Fabiano, Department of Counseling, School, and Educational Psychology, University at Buffalo, State University of New York.

The manuscript is based on data used for Brian Daniels' doctoral dissertation. The screener evaluated in the study was published in a recent book by Robert J. Volpe and Gregory A. Fabiano, entitled *Daily Behavior Report Cards: An Evidence-Based System of Assessment and Intervention*. Drs. Volpe and Fabiano receive royalties from the publisher for sales of the book.

Correspondence concerning this article should be addressed to Brian Daniels, Department of Counseling and Applied Educational Psychology, Northeastern University, Boston, MA 02115. E-mail: daniels.br@husky.neu.edu

measure to be appropriate for school-based screening, it should provide timely and useful information regarding levels of risk for all students, and it must be compatible with available services.

For a screening measure to be technically adequate, it should demonstrate acceptable reliability, validity and classification accuracy. Reliability refers to the degree to which scores are consistent across administrations. Validity refers to the degree to which there is evidence to support the use of the results for an intended purpose. Classification accuracy refers to the extent to which a screening measure effectively differentiates at-risk children from typically developing children.

For a screening measure to be usable, it should be feasible and acceptable to stakeholders, and data obtained from the screening measure should identify at-risk students, guide selection of interventions, inform goals for intervention, and complement progress monitoring assessments (Dowdy, Furlong, Eklund, Saeki, & Ritchey, 2010; Glover & Albers, 2007). When school-wide screening is performed, it is not always clear what should be done with the resultant data; therefore, consideration should be given to how screening results may inform intervention in designing screening protocols. Indeed, intervention planning was rated by school psychologists as the most frequent and useful assessment purpose (Ogg et al., 2013).

Extant Behavioral Screening Methods

Severson, Walker, Hope-Doolittle, Kratochwill, and Gresham (2007) outlined the following three methods for conducting school-based behavioral screening: (a) multiple-gating procedures (e.g., Systematic Screening for Behavior Disorders [SSBD; Walker & Severson, 1990, 1992]), (b) teacher evaluation of all students in a classroom with regard to common behavioral criteria (e.g., Student Risk Screening Scale [SRSS; Drummond, 1994]), and (c) teacher nomination of problem students followed by individual ratings of nominated students using a comprehensive rating scale (e.g., BASC-2 Behavioral and Emotional Screening System [BESS; Kamphaus & Reynolds, 2007]). Although all have demonstrated technical adequacy, the three methods differ with regard to their feasibility. Individual teacher-report ratings require significant time, and multiple-gating procedures typically require significant school resources, which may limit their feasibility for school-based screening. The usability of the three methods is also limited in that all three have been used solely for the purpose of classification (i.e., identifying at-risk students). Item analysis may be used with any of these methods to identify specific problem behaviors, but items typically represent problem behaviors that are not explicitly linked to treatment and may not be appropriate targets for treatment (e.g., "fidgets") (Gresham, 2005; Pelham et al., 2005). Indeed, assessment for screening and assessment for treatment planning are tasks that have been siloed in school-based practice (Cook, Volpe, & Livanis, 2010), with the former used to identify global problems and the latter reserved to identify solutions in those whom have screened positive.

Integrated Screening and Intervention

Volpe, Briesch, and Chafouleas (2010) conceptualized an adaptive model of behavior assessment that uses screening results to inform behavioral goals for intervention and to generate brief, feasible progress monitoring measures. This model served as the foundation for the Integrated Screening and Intervention System (ISIS; Volpe & Fabiano, 2013), which was designed to identify and treat students with classroom behavior problems commonly experienced by students with ADHD. The ISIS employs a multiple-gating screening procedure with three distinct improvements over extant screening methods: (a) screening focuses on specific problem behaviors impacting classroom functioning as opposed to diagnostic symptoms for the purpose of classification, (b) screening generates specific target behaviors that may become the focus of an evidence-based intervention, and (c) screening directly informs the development of a feasible method for monitoring student progress in the targeted areas.

Screening in the ISIS model involves teacher nomination and rating of five students in a given classroom whom exhibit concerning behaviors using the ISIS Teacher Rating Form (ITRF; Volpe & Fabiano, 2013). The 43 items of the ITRF are based upon the most commonly used Daily Behavior Report Card (DRC) items employed in a study evaluating the effectiveness of DRC for students with ADHD (Fabiano et al., 2010). A conversion table links each ITRF item back to the DRC item(s) from which it was derived. The ITRF was designed so that screening data could be used: (a) to quantify the level of risk for each rated student, (b) to rank order the level of risk for students within classrooms, and (c) to rank order problem behaviors to aid in the selection of target behaviors for DRC intervention. A screening instrument must differentiate between related but independent domains within a construct of interest to be sensitive enough to quantify and prioritize risk among students who exhibit problems in only one domain (Hinshaw, 1987). For example, students who are at risk for attention problems alone may not reach a screening criterion based on a global scale (combined attention and conduct problems) and would therefore not be identified for intervention.

DRC is an intervention wherein students may earn rewards for reaching predetermined criteria for clearly defined behaviors that are rated by teachers one or more times per school day (see Volpe & Fabiano, 2013). Research has shown DRC to be an efficient and effective intervention for problem behaviors exhibited by students (Chafouleas, Riley-Tillman, & McDougal, 2002; Vannest, Davis, Davis, Mason, & Burke, 2010). Also, DRC themselves can be used as progress monitoring measures (Fabiano, Vujnovic, Naylor, Pariseau, & Robins, 2009; Pelham et al., 2005).

Purpose

An integrated system of screening and intervention has the potential to improve upon existing screening measures by balancing feasibility with sound psychometric properties and treatment validity, or the degree to which assessment practices result in beneficial outcomes (Cone, 1988; Hayes, Nelson, & Jarrett, 1987). The purpose of the study is to investigate the factor structure, reliability and validity of the ITRF. The technical characteristics of the ITRF have not been previously evaluated; therefore, no data regarding the factor structure, reliability or validity were available before the study. Identification of the factor structure is essential to validate the ITRF for detecting problems in each of the distinct attention- and conduct-problem domains within the overarching construct of ADHD (Hinshaw, 1987; Loney & Milich, 1982). The following research questions guided this study:

1. What is the factor structure of the ITRF?
2. Does the ITRF demonstrate adequate internal consistency and temporal stability?
3. Do the patterns of associations between the scales of the ITRF and relevant scales from a widely used, commercially available rating scale support the convergent and discriminant validity of the ITRF?

Method

Participants and Setting

A total of 39 teachers from five elementary schools in two suburban school districts in the northeastern United States participated in the study. All participants were general education classroom teachers (grades K-6). First, in a manner consistent with the intended purpose of the ITRF, each teacher nominated the five students in his or her class who most exhibited behaviors that interfered with learning and completed ITRF ratings for these five students. Next, each teacher was asked to complete a second ITRF for five different students randomly selected from his or her class roster.

Teachers completed ratings for 390 students (240 male; 150 female) in grades K-6. Males comprised a larger proportion of the nominated students (75.4% male; 24.6% female) than the overall sample (61.5% male; 38.5% female), which is consistent with prior research involving teacher nomination of students with behavior problems (Soles, Bloom, Heath, & Karagiannakis, 2008). Table 1 summarizes the age and grade distributions of students for whom teacher ratings were obtained. Because all student information was provided by teachers, official socioeconomic data were not available; however, percentages of students identified as receiving free and reduced-priced lunch in the five participating schools were as follows: 7.9%, 10.4%, 13.5%, 19.8% and 22.2%. Table 2 compares student ethnicity across the total sample, nominated and randomly selected subsamples, and five schools from which data were collected.

Table 1
Distribution of Students by Age and Grade

	Frequency	Percent
Age		
5	1	0.3
6	29	7.4
7	31	7.9
8	61	15.6
9	78	20.0
10	100	25.6
11	70	17.9
12	17	4.4
13	3	0.8
Grade		
K	20	5.1
1	30	7.7
2	60	15.4
3	80	20.5
4	90	23.1
5	90	23.1
6	20	5.1

Analysis of demographic data indicated a higher percentage of African American students was nominated by teachers than would be expected based on the percentages of African Americans in schools from which data were collected. Prior studies indicate teacher ratings of problem behavior are often elevated for African American students (Epstein, March, Conners, & Jackson, 1998; Reid, Casat, Norton, Anastopoulos, & Temple, 2001); therefore, it is possible that the higher representation of African American students in the nominated subsample was the result of teacher bias.

Procedures

Teachers completed all behavioral measures (described below) while the primary researcher was present to verify all measures were accurately completed in accordance with instructions. Teachers required between 5 and 10 minutes to complete each ITRF for five students. A subset of 26 teachers individually completed ITRF ratings for the same students 2–4 weeks following the first rating, which resulted in 260 student ratings (130 nominated and 130 randomly selected) for temporal stability analyses. A 2-week target interval was selected to reduce the influence of maturation on teacher ratings (Nunnally & Bernstein, 1994).

Measures

Integrated Screening and Intervention System Teacher Rating Form. The ISIS Teacher Rating Form (ITRF; Volpe & Fabiano, 2013) was developed as a screening tool to identify and rank order students who exhibit problem behaviors that may interfere with learning. The 43 items on the ITRF were derived from target behaviors frequently used in a DRC intervention study for students with ADHD (Fabiano et al., 2010), and the ITRF items were further refined based on the results of a pilot study (Volpe & Fabiano, 2013).

The first step in the ITRF screening process requires teachers to identify five students in their classes who exhibit behaviors that most "Interfere with their learning or the learning of others" (Volpe & Fabiano, 2013). Next, teachers rank the five identified students starting with

Table 2
Ethnic Composition of the Total Sample and Subsamples

	Mean (5 Schools)	Total sample (N = 390)	Nominated (n = 195)	Randomly selected (n = 195)
White	86.4%	85.4%	84.1%	86.7%
African American	2.4%	4.1%	6.7%	1.5%
Latino	3.3%	2.6%	1.5%	3.6%
Asian	4.8%	5.4%	3.6%	7.2%
Multirace/Other	3.0%	2.6%	4.1%	1.0%

Note. The Asian category includes people having origins in the Far East, Southeast Asia, the Indian subcontinent, Hawaii, Guam, Samoa, or other Pacific Islands. Mean percentages were calculated from individual school data obtained from the state department of education. Total Sample, Nominated subsample, and Randomly Selected subsample data were reported by classroom teachers.

the student who exhibits the most concerning behaviors. Finally, teachers read each item of the ITRF and indicate their level of concern with regard to each student on a four-point scale (1 = *slight concern*, 2 = *moderate concern*, 3 = *strong concern*, and the space is left blank if the behavior is not a concern). The matrix format of the ITRF allows teachers to complete simultaneous ratings for all identified students one item at a time. Teachers only need to read each item once to complete ratings for all five students, which reduces the amount of time necessary for screening when compared with individual rating scales that require teachers to reread each item for every student. Item ratings are summed for each student to generate a total score, which prioritizes students for intervention, and items with the highest ratings may be converted to potential targets for DRC interventions.

Brief Problem Monitor-Teacher. The Brief Problem Monitor-Teacher (BPM-T; Achenbach, McConaughy, Ivanova, & Rescorla, 2011) is an 18-item empirically derived teacher-report rating scale developed for monitoring the effects of intervention. Teachers respond to each item using a three-point scale (0 = *not true*, 1 = *somewhat true*, and 2 = *very true*). Separate norms are available for males and females that yield normalized *T* scores for Total Problems, Internalizing (6 items), Externalizing (6 items), and Attention Problems (6 items) scales based on a U.S. national sample. *T* scores ≥65 are considered elevated and indicate concern with regard to the respective problem area(s), whereas *T* scores <65 are considered to be in the normal range.

Achenbach et al. (2011) reported coefficients of stability ranging from .86 to .93 across a 16-day interval, and alpha coefficients ranging from .80 to .90 for the BPM-T. Evidence for criterion-related validity was provided by multiple regression analyses of BPM-T scale scores for children referred for mental health services versus nonreferred children. BPM-T scale scores were significantly higher for referred than nonreferred children, and effect sizes, which reflected the percentages of variance in BPM-*T* scores accounted for by referral status when controlling for age, socioeconomic status, and ethnicity, ranged from 16% to 29% for the BPM-T (Achenbach et al., 2011).

Data Analysis

Exploratory factor analysis (EFA) was performed to identify latent variables that accounted for the pattern of correlations among manifest variables (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Initial data screening indicated the presence of numerous multivariate outliers; therefore, principal axis factoring (PAF) was used to extract factors because it does not require multivariate normality (Fabrigar et al., 1999). The number of factors identified for retention was initially guided by analysis of the scree plot and parallel analysis (PA; Horn, 1965). PA was conducted by computing eigenvalues from multiple sample correlation matrices generated from permutations of raw data. Hypothetically, factors would be retained if their values exceeded the 95th percentile of eigenvalues generated from the permuted data. Although factor retention decisions were initially guided by scree plot data and the results of PA, item retention criteria were ultimately applied to ensure the number of identified factors would result in subscales comprised of at least four items per subscale (Russell, 2002). Factor solutions were systematically evaluated with decreasing numbers of factors, beginning with the number of factors identified by PA, until a factor solution resulting in viable subscales comprised of at least four items each was identified. Finally, direct quartimin rotation (an oblique method) was applied because the factors generated by the ITRF were expected to be correlated with one another (Costello & Osborne, 2005; Fabrigar & Wegener, 2012).

An item was considered for inclusion in an ITRF subscale if the absolute factor loading was greater than or equal .32 (Tabachnick & Fidell, 2001). However, an item was excluded from subscales if it demonstrated cross-loadings within .15 of the item's highest factor loading or if the absolute value of factor loadings was higher than .32 on two or more factors (Worthington & Whittaker, 2006). Item communalities following rotation were also examined because they represent the proportion of item variance accounted for by the factors (Tabachnick & Fidell, 2001).

Cronbach's alpha greater than or equal to .80 was used as the criterion to determine adequate internal consistency (Cicchetti, 1994; Nunnally, & Bernstein, 1994). Pearson product–moment correlation coefficients were calculated to assess the temporal stability, convergent validity and discriminant validity of the ITRF. Correlation coefficients greater than or equal to .80 for the ITRF total and subscale scores across 2– to 4-week intervals were considered indicative of good temporal stability (Bracken, 1987). Paired-samples *t* tests were also conducted to determine if ITRF total and subscale scores differed across 2–4 weeks. To examine convergent validity, we examined Pearson product–moment correlations between each factor extracted from the ITRF and the most relevant corresponding scale from the BPM-T. Discriminant validity was assessed by calculating Pearson product–moment correlations between ITRF subscales and dissimilar scales from the BPM-T. Cohen's (1988) guidelines provided standards for evaluating the magnitude of small ($r = .10$), medium ($r = .30$) and large ($r = .50$) correlation coefficients between the measures.

Results

Research Question 1: What Is the Factor Structure of the ITRF?

Missing values analysis did not reveal any missing data. Distributions of teacher ratings were skewed and kurtotic for every ITRF item and several multivariate outliers were identified by comparing obtained values with critical values of Mahalanobis distance, χ^2 (40) = 73.402, $p < .001$. Inverse transformation was performed on all ITRF variables to reduce skewness, kurtosis and the influence of multivariate outliers.

Before performing EFA, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling adequacy were reviewed. Bartlett's test of sphericity was significant ($p < .001$), which indicated adequate relationships between variables (i.e., the correlation matrix was not an identity matrix), and the KMO index reached an acceptable value (.954), which indicated the absence of multicollinearity.

EFA indicated all ITRF items substantially loaded on a single factor (total score) in the unrotated solution. The single factor accounted for 41.63% of the variance in teacher ratings, and item loadings on the first factor ranged from .49 to .76.

Results of PA initially indicated six factors for retention and are reported in Table 3. In contrast, visual analysis of the scree plot indicated a clear break at the third factor; however, numerous items with low factor loadings or substantial loadings on more than one factor were identified. Subjecting items to aforementioned inclusion criteria would have resulted in insufficient items to form subscales comprised of at least four items each from the sixth, fifth, fourth and third factors (Fabrigar et al., 1999; Russell, 2002). Because the primary purpose of the ITRF is to inform subsequent intervention by prioritizing problem behaviors that guide the selection of target behaviors for intervention, the exclusion of multiple items from potential subscales was deemed to be excessive. Therefore, a two-factor solution was selected to include as many ITRF items as possible in the development of subscales comprised of at least four items each. Although item communalities were low (e.g., < .50), adequate recovery of the factor structure was still possible because there was a small number of overdetermined factors (e.g., more than seven items with substantial loadings per factor) and the sample was relatively large (MacCallum, Widaman, Zhang, & Hong, 1999).

The total variance accounted for by the first six factors is provided in Table 4. The two factors in the interpreted solution accounted for 49.52% of the cumulative variance in teacher ratings. Pattern and structure coefficients for ITRF items in the rotated two-factor solution are reported in Table 5. The two factors were

Table 3
Results of Parallel Analysis: Raw and Permuted Data Eigenvalues

		Permuted data	
Root	Raw data	Mean	95th percentile
1	17.94	0.82	0.90
2	3.43	0.74	0.80
3	1.73	0.68	0.73
4	1.14	0.63	0.68
5	0.93	0.58	0.63
6	0.73	0.54	0.58
7	0.53	0.50	0.54

Note. N = 390; Number of Variables = 43; Number of Datasets = 5,000.

Table 4
Total Variance Explained

Factor	Extraction sums of squared loadings Total	Percentage of variance	Cumulative percentage
1	17.90	41.63	41.63
2	3.40	7.90	49.52
3	1.69	3.93	53.45
4	1.12	2.61	56.06
5	0.88	2.05	58.11
6	0.70	1.62	59.72

named Oppositional/Disruptive (25 items) and Academic Productivity/Disorganization (16 items). The following two items had loadings greater than .32 on both factors, which violated a priori subscale inclusion criteria: *transitions poorly between activities* and *takes too long when using bathroom or water fountain*. As a result, the two items were not included in subscales used in subsequent analyses, although they were retained in the ITRF total score given the utility of the items for informing DRC intervention.

Research Question 2: Does the ITRF Demonstrate Adequate Internal Consistency and Temporal Stability?

Results indicated acceptable internal consistency ($\alpha \geq .80$) of the ITRF total score with alpha coefficients equal to .97 for both ITRF administrations. Alpha coefficients for the Oppositional/Disruptive subscale were .95 and .96 for the first and second administrations, respectively. Alpha coefficients for the Academic Productivity/Disorganization subscale were .94 and .95 for the first and second administrations, respectively.

Coefficients of stability were calculated for 67% of the sample across intervals ranging from 2–4 weeks. Coefficients of stability were found to be good for the ITRF total score ($r = .84$) and Academic Productivity/Disorganization subscale ($r = .88$). The coefficient of stability for the Oppositional/Disruptive subscale approached an acceptable level ($r = .78$). Paired-samples t tests of nominated student ratings using inversely transformed data indicated significantly lower mean ITRF total, $t(129) = -3.26$, $p = .001$, $d = -0.29$, 95% CI [−0.04, −0.01], Oppositional/Disruptive, $t(129) = -3.40$, $p = .001$, $d = -0.30$, 95% CI [−0.10, −0.03], and Academic Productivity/Disorganization scores, $t(129) = -3.41$, $p = .001$, $d = -0.30$, 95% CI [−0.04, −0.01] after 2–4 weeks. Paired-samples t tests of randomly selected student ratings using inversely transformed data indicated a significantly higher mean ITRF total score, $t(129) = -2.03$, $p = .045$, $d = -0.18$, 95% CI [−0.08, −0.001] after 2–4 weeks. Oppositional/Disruptive, $t(129) = -0.21$, $p = .837$, $d = -0.02$, 95% CI [−0.05, 0.04], and Academic Productivity/Disorganization scores, $t(129) = -1.06$, $p = .289$, $d = -0.09$, 95% CI [−0.08, −0.02] were not significantly different for randomly selected students after 2–4 weeks.

Research Question 3: Do the Patterns of Associations Between the Scales of the ITRF and Relevant Scales From a Widely Used, Commercially Available Rating Scale Support the Convergent and Discriminant Validity of the ITRF?

Mean ITRF and BPM-T ratings for the total sample and nominated and randomly selected subsamples at Time 1 are reported in Table 6. Correlation coefficients between ITRF and BPM-T scores are summarized in Table 7. Evidence for the convergent validity of the ITRF was indicated by large correlation coefficients between the ITRF total score and the BPM-T Total Problems score ($r = .81$) and between the ITRF subscales and the BPM-T Externalizing and Attention Problems scales (rs = .52–.80). As expected, correlation coefficients were highest between the ITRF Oppositional/Disruptive subscale and the BPM-T Externalizing scale ($r = .80$) and between the ITRF Academic Productivity/Disorganization subscale and the BPM-T Attention Problems scale ($r = .76$). Evidence for discriminant validity was provided by smaller correlation coefficients between distinct constructs including the ITRF Oppositional/Disruptive subscale and the BPM-T Attention Problem scale ($r = .58$) and the ITRF Academic Productivity/Disorganization subscale and the BPM-T Externalizing scale ($r = .52$). Smaller correlation coefficients were also found between dissimilar constructs such as the ITRF subscales and the BPM-T Internalizing scale (rs = .34–.40).

Table 5
Pattern and Structure Matrices

	Pattern coefficients		Structure coefficients	
	Factor 1	Factor 2	Factor 1	Factor 2
Argues with peers[a]	.84		.76	.34
Argues with teacher[a]	.79		.79	.46
Disrespectful to adults[a]	.77		.80	.50
Calls out[a]	.76		.71	.35
Bossy[a]	.76		.61	
Has conflicts with peers[a]	.74		.74	.43
Loses temper[a]	.74		.75	.44
Bumps, hits or kicks others[a]	.72		.73	.42
Tattles on other children[a]	.72		.68	.34
Uses inappropriate language[a]	.69		.71	.43
Noncompliant[a]	.69		.75	.51
Teases[a]	.69		.66	.35
Does not respect others' personal space[a]	.68		.75	.51
Does not work well with others[a]	.65		.74	.52
Nosey[a]	.64		.60	
Uses materials inappropriately[a]	.64		.75	.56
Disrupts others[a]	.63		.72	.51
Makes irrelevant comments[a]	.60		.67	.47
Destroys property[a]	.59		.66	.47
Moves around the room[a]	.57		.65	.48
Cries, complains, whines[a]	.52		.61	.45
Distracted by others' negative behaviors[a]	.51		.61	.47
Leaves room without permission[a]	.51		.60	.44
Asks to leave classroom frequently[a]	.47		.62	.53
Makes self-deprecating comments[a]	.45		.56	.45
Does not turn in class assignments[b]		.89	.41	.83
Unorganized[b]		.82	.40	.78
Missing or incomplete homework[b]		.80	.37	.75
Does not complete classwork on time[b]		.80	.39	.75
Comes to class unprepared[b]		.77	.38	.73
Fails to pack needed materials for home[b]		.75	.43	.75
Inaccurate or incomplete classwork[b]		.72	.46	.75
Does not start assignments independently[b]		.71	.48	.75
Does not correct own work[b]		.59	.49	.68
Does not put away belongings[b]		.58	.48	.67
Does not participate in class[b]		.57	.42	.62
Does not ask for help/asks for help inappropriately[b]		.52	.54	.67
Does not follow directions[b]		.50	.59	.68
Writes illegibly[b]		.50	.43	.58
Does not participate in group activities[b]		.43	.45	.54
Mumbles or speaks incoherently[b]		.40	.45	.52
Transitions poorly between activities	.43	.33	.62	.58
Takes too long when using bathroom or water fountain	.37	.34	.57	.56

Note. Coefficients less than .30 were suppressed.
[a] Items retained within the *ISIS Teacher Rating Form* (ITRF) Oppositional/Disruptive subscale. [b] Items retained within the ITRF Academic Productivity/Disorganization subscale.

Discussion

The purpose of the study was to investigate the factor structure, reliability and validity of a novel teacher-completed screening instrument, which is directly linked to an evidence-based intervention. EFA yielded a two-factor solution with factors encompassing oppositional or disruptive behaviors and academic productivity problems. Although the majority of

Table 6
Mean ITRF and BPM-T Scores and Standard Deviations at Time 1

	Total sample (N = 390)	Nominated (n = 195)	Randomly-selected (n = 195)
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
ISIS Teacher Rating Form			
Oppositional/Disruptive	12.5 (15.2)	20.4 (16.7)	4.7 (7.7)
Academic Productivity/Disorganization	13.3 (12.5)	20.9 (11.7)	5.7 (7.9)
Total Score	27.1 (26.4)	43.4 (25.9)	10.8 (13.9)
ASEBA Brief Problem Monitor-Teacher			
Externalizing	1.5 (2.4)	2.6 (2.8)	0.4 (0.9)
Attention	4.7 (3.9)	7.4 (3.2)	2.0 (2.6)
Internalizing	2.4 (2.9)	3.4 (3.3)	1.4 (2.0)
Total Problems	8.6 (7.2)	13.4 (6.4)	3.8 (3.9)

ITRF items demonstrated simple loadings (loading on only one factor), two items had substantial loadings (e.g., ≥ .32) on both factors and were excluded from the subscales; however, the items remained in the calculation of the ITRF total score because of their usefulness for informing intervention targets. Results indicated adequate internal consistency and convergent validity of the ITRF total score and Academic Productivity/Disorganization and Oppositional/Disruptive subscale scores. Evidence for temporal stability was mixed. Coefficients of stability were acceptable for the ITRF total score and Academic Productivity/Disorganization subscale, and although the stability coefficient estimate for the Oppositional/Disruptive subscale was slightly lower, it exceeded the minimum standard recommended by some authors (Kline, 2000). In contrast to correlational analyses, results of paired-samples *t* tests revealed significantly lower mean ITRF Oppositional/Disruptive, Academic Productivity/Disorganization, and total scores for nominated students, and a significantly higher mean ITRF total score for randomly selected students after 2–4 weeks. Although interventions were not formally implemented for nominated students in the study, it is possible that individual teachers developed and implemented interventions on their own following the first rating, which may have accounted for

Table 7
Pearson Correlation Coefficients Between ITRF Subscales and BPM-T Scales

	ISIS Teacher Rating Form			ASEBA Brief Problem Monitor-Teacher			
	Oppositional/ Disruptive	Academic Productivity/ Disorganization	Total score	Externalizing	Attention	Internalizing	Total problems
ISIS Teacher Rating Form							
Oppositional/Disruptive	—	.64*	.92*	.80*	.58*	.34*	.72*
Academic Productivity/ Disorganization		—	.89*	.52*	.76*	.40*	.75*
Total Score			—	.74*	.73*	.40*	.81*
ASEBA Brief Problem Monitor-Teacher							
Externalizing				—	.47*	.35*	.74*
Attention					—	.34*	.85*
Internalizing						—	.71*
Total Problems							—

* $p < .001$.

lower ratings of problem behavior several weeks later. However, lower ratings of problem behavior for nominated students and higher overall ratings of problem behavior for randomly selected students may potentially be explained by regression toward the mean. Effect sizes indicated small changes in teacher ratings across 2–4 weeks that may have resulted from less extreme ratings of student behavior over time (i.e., behaviors were rated as less severe for nominated students and more typical for randomly selected students). It is worth noting that the items of the ITRF were designed to assess classroom impairment as opposed to psychiatric symptoms. As such, it would be expected that scores on the ITRF would be less stable than scores on rating forms designed solely for classification purposes. Nevertheless, coefficients of stability reached or approximated acceptable levels, indicating that students tended to maintain their rank among peers over time.

Overall, ITRF subscales appeared to measure domains that encompass behaviors associated with ADHD. First, academic underachievement is a key domain of functional impairment in students with ADHD (Frazier, Youngstrom, Glutting, & Watkins, 2007; McConaughy et al., 2011). Second, the relationship between ADHD and oppositional behavior is well documented (see Barkley, 2006), and elevated oppositional behavior may be a useful indicator of treatment resistance and the emergence of more severe behavior problems (e.g., conduct disorder) over time (e.g., Kolko & Pardini, 2010; Loeber, Burke, & Pardini, 2009). The two-factor structure of the ITRF allows stakeholders to differentiate between ADHD domains to quantify and prioritize risk among students who may only exhibit problems in one domain and who may not be identified as at risk when evaluated against an overall total score and threshold based on combined attention and conduct problems.

Although EFA did not yield a distinct interpersonal problems factor, the ITRF includes items that assess social functioning with regard to peers (e.g., *bumps, hits or kicks others*, *does not respect others' personal space*, *teases*, *has conflicts with peers*, *argues with peers*, and *does not work well with others*), which may be attributable to oppositional and/or impulse-control problems.

Several screening tools currently exist to assess symptoms of ADHD; however, the ITRF differs in that it assesses classroom behaviors that affect student learning. The focus on problem behaviors over underlying psychopathology gives the ITRF two distinct advantages over other ADHD assessments: (a) the ITRF allows school psychologists and teachers to use a preventive approach to identifying children who may be at risk for academic, social or behavioral problems without labeling or stigmatizing students in order to provide intervention before significant problems develop and (b) the ITRF directly informs intervention for behaviors associated with externalizing and academic performance problems by linking behaviors that interfere with student learning to specific targets that may be included on a DRC for both intervention and progress monitoring. For example, items on the ITRF such as *calls out* and *moves around the room* correspond to DRC items intended to increase the frequency of corresponding target behaviors such as *raises hand to speak with X or fewer reminders* and *sits appropriately in assigned area with X or fewer reminders*, respectively. Finally, the feasibility of the ITRF gives the ITRF an advantage over other extant screening tools. Teacher nomination and rating using the ITRF required only 5–10 minutes, which is considerably less than the 1–1.5 hours required for teacher nomination and rating in SSBD stages 1 and 2.

Although the ITRF identifies students who may benefit from intervention, it represents only the first stage of the ISIS model. Following completion of the ITRF, the school psychologist may meet with teachers, individually or as a group, to help them identify the three to five most concerning problem behaviors for each student indicated by ITRF ratings. DRC intervention design is facilitated by a conversion table, which directly links ITRF items to target behaviors. Classroom teachers subsequently record baseline data for one week to determine appropriate target criteria for DRC items. Finally, DRC interventions are implemented with home-based contingencies and students' daily performance is monitored. Once a student achieves his or her goal for a particular target behavior, it is replaced on the DRC by another high-priority behavior identified by ITRF ratings.

The results of the study provide initial evidence for the technical adequacy of the ITRF, but there are several limitations that warrant consideration. First, data were collected via teacher report with each teacher conducting ratings for multiple students in his or her class. It is possible that nested data may result in inflated factor loadings and overestimation of the variance accounted for by the factors (Julian, 2001). A substantial amount of the explained variance may be attributable to the influence of each student's behavior on the behavior of other students in the classroom or to differences among classroom environments, including differences in teachers' classroom management skills, attitudes and tolerances for disruptive behavior rather than by actual differences in student behavior. Although the ITRF asks teachers to complete ratings for five students, future studies may seek to include a larger sample of teachers who each complete ratings for one student to provide truly independent observations that may be subjected to factor and criterion-related validity analyses. Future studies should also seek to verify the two-factor structure of the ITRF through confirmatory factor analysis using a different sample.

Second, the study used a relatively homogeneous sample comprised primarily of White children, which was geographically restricted to the northeastern United States; therefore, the results of the study may not generalize to other populations. It is possible that teacher ratings in the study reflected behavioral expectations for nonminority students in suburban schools, which resulted in biased assessment of minority students' behavior. It appears that our results are consistent with those of prior studies that indicated elevated ratings for African American students' problem behavior (Epstein et al., 1998; Reid et al., 2001); however, additional studies are needed to determine whether such differences are because of informant bias or whether such differences also would be found using more objective assessment methods (see Hosterman, DuPaul, & Jitendra, 2008). Finally, future research should investigate the feasibility of the ITRF for large-scale universal screening by comparing the time and resources necessary to conduct universal screening across an entire school or district using the ITRF with those necessary to conduct screening using extant measures. Acceptability may be assessed by gathering teachers' attitudes and perceptions regarding the ITRF after using the full ISIS model in their classrooms for students who exhibit disruptive behaviors or academic productivity problems. In closing, the ITRF holds promise for facilitating school-based screening, evidence-based intervention, and progress monitoring of behaviors associated with ADHD.

References

Achenbach, T. M., McConaughy, S. H., Ivanova, M. Y., & Rescorla, L. A. (2011). *Manual for the ASEBA Brief Problem Monitor*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2007). Long-term school outcomes for children with attention-deficit/hyperactivity disorder: A population-based perspective. *Journal of Developmental and Behavioral Pediatrics, 28,* 265–273. http://dx.doi.org/10.1097/DBP.0b013e31811ff87d

Barkley, R. A. (2006). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). New York, NY: Guilford Press.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 4,* 313–326. http://dx.doi.org/10.1177/073428298700500402

Chafouleas, S. M., Riley-Tillman, T. C., & McDougal, J. L. (2002). Good, bad, or in-between: How does the daily behavior report card rate? *Psychology in the Schools, 39,* 157–169. http://dx.doi.org/10.1002/pits.10027

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290. http://dx.doi.org/10.1037/1040-3590.6.4.284

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

Cone, J. D. (1988). Psychometric considerations and multiple models of behavioral assessment. In A. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 42–66). New York, NY: Pergamon Press.

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment* *for Effective Intervention, 35,* 197–205. http://dx.doi.org/10.1177/1534508410379842

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*. Retrieved from http://pareonline.net/pdf/v10n7.pdf

Dowdy, E. T., Furlong, M. J., Eklund, K., Saeki, E., & Ritchey, K. (2010). Screening for mental health and wellness: Current school-based practices and emerging possibilities. In B. Doll, W. Pfohl, & J. Yoon (Eds.), *Handbook of youth prevention science* (pp. 70–95). New York, NY: Routledge.

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

DuPaul, G. J. (2004). ADHD identification and assessment: Basic guidelines for educators. *Helping children at home and school II: Handouts for families and educators*. Bethesda, MD: National Association of School Psychologists.

DuPaul, G. J., & Kern, L. (2011). *Young children with ADHD: Early identification and intervention*. Washington, DC: American Psychological Association.

Dwyer, S. B., Nicholson, J. M., & Battistutta, D. (2006). Parent and teacher identification of children at risk of developing internalizing or externalizing mental health problems: A comparison of screening methods. *Prevention Science, 7,* 343–357. http://dx.doi.org/10.1007/s11121-006-0026-5

Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology, 26,* 109–118. http://dx.doi.org/10.1023/A:1022617821422

Fabiano, G. A., Vujnovic, R., Naylor, J., Pariseau, M., & Robins, M. (2009). An investigation of the technical adequacy of a daily behavior report card (DBRC) for monitoring progress of students with attention-deficit/hyperactivity disorder in special education placements. *Assessment for Effective Intervention, 34,* 231–241. http://dx.doi.org/10.1177/1534508409333344

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

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York, NY: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299. http://dx.doi.org/10.1037/1082-989X.4.3.272

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: Meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities, 40,* 49–65. http://dx.doi.org/10.1177/00222194070400010401

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

Gresham, F. M. (2005). Response to intervention: An alternative means to identifying students as emotionally disturbed. *Education & Treatment of Children, 28,* 328–344.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment. A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963–974. http://dx.doi.org/10.1037/0003-066X.42.11.963

Hinshaw, S. P. (1987). On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin, 101,* 443–463. http://dx.doi.org/10.1037/0033-2909.101.3.443

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. http://dx.doi.org/10.1007/BF02289447

Hosterman, S. J., DuPaul, G. J., & Jitendra, A. K. (2008). Teacher ratings of ADHD symptoms in ethnic minority students: Bias or behavioral difference? *School Psychology Quarterly, 23,* 418–435. http://dx.doi.org/10.1037/a0012668

Jones, D., Dodge, K. A., Foster, E. M., Nix, R., & the Conduct Problems Prevention Research Group. (2002). Early identification of children at risk for costly mental health service use. *Prevention Science, 3,* 247–256. http://dx.doi.org/10.1023/A:1020896607298

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8,* 325–352. http://dx.doi.org/10.1207/S15328007SEM0803_1

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31–36. http://dx.doi.org/10.1007/BF02291575

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

Kline, P. (2000). *Handbook of psychological testing*. New York, NY: Routledge.

Kolko, D. J., & Pardini, D. A. (2010). ODD dimensions, ADHD, and callous-unemotional traits as predictors of treatment response in children with disruptive behavior disorders. *Journal of Abnormal Psychology, 119,* 713–725. http://dx.doi.org/10.1037/a0020910

Loeber, R., Burke, J., & Pardini, D. A. (2009). Perspectives on oppositional defiant disorder, conduct disorder, and psychopathic features. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 50,* 133–142. http://dx.doi.org/10.1111/j.1469-7610.2008.02011.x

Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich & D. K. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 3, pp. 113–147). Greenwich, CT: JAI.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99. http://dx.doi.org/10.1037/1082-989X.4.1.84

McConaughy, S. H., Volpe, R. J., Antshel, K. M., Gordon, M., & Eiraldi, R. B. (2011). Academic and social impairments of elementary school children with attention deficit hyperactivity disorder. *School Psychology Review, 40,* 200–225.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Ogg, J., Fefer, S., Sundman-Wheat, A., McMahan, M., Stewart, T., Chappel, A., & Bateman, L. (2013). School-based assessment of ADHD: Purpose, alignment with best practice guidelines, and training. *Journal of Applied School Psychology, 29,* 305–327. http://dx.doi.org/10.1080/15377903.2013.836775

Pelham, W. E., Burrows-Maclean, L., Gnagy, E. M., Fabiano, G. A., Coles, E. K., Tresco, K. E., . . . Hoffman, M. T. (2005). Transdermal methylphenidate, behavioral, and combined treatment for children with ADHD. *Experimental and Clinical Psychopharmacology, 13,* 111–126. http://dx.doi.org/10.1037/1064-1297.13.2.111

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

Reid, R., Casat, C. D., Norton, H. J., Anastopoulos, A. D., & Temple, E. P. (2001). Using behavior rating scales for ADHD across ethnic groups: The IOWA Conners. *Journal of Emotional and Behavioral Disorders, 9,* 210–218. http://dx.doi.org/10.1177/106342660100900401

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28,* 1629–1646. http://dx.doi.org/10.1177/014616702237645

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. http://dx.doi.org/10.1016/j.jsp.2006.11.003

Soles, T., Bloom, E. L., Heath, N. L., & Karagiannakis, A. (2008). An exploration of teachers' current perceptions of children with emotional and behavioural difficulties. *Emotional & Behavioural Difficulties, 13,* 275–290. http://dx.doi.org/10.1080/13632750802442201

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Tobin, R. M., Schneider, W. J., Reck, S., & Landau, S. (2008). Best practices in the assessment of children with attention deficit hyperactivity disorder: Linking assessment to response to intervention. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-V* (pp. 617–631). Bethesda, MD: National Association of School Psychologists.

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review, 39,* 654–672.

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., . . . Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child and Adolescent Psychiatry, 53,* 34–46.e2. http://dx.doi.org/10.1016/j.jaac.2013.09.001

Visser, S. N., Lesesne, C. A., & Perou, R. (2007). National estimates and factors associated with medication treatment for childhood attention-deficit/hyperactivity disorder. *Pediatrics, 119,* S99–S106. http://dx.doi.org/10.1542/peds.2006-2089O

Volpe, R. J., Briesch, A. M., & Chafouleas, S. M. (2010). Linking screening for emotional and behavioral problems to problem-solving efforts: An adaptive model of behavioral assessment. *Assessment for Effective Intervention, 35,* 240–244. http://dx.doi.org/10.1177/1534508410377194

Volpe, R. J., & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of assessment and intervention*. New York, NY: Guilford Press.

Walker, H. M., & Severson, H. H. (1990). *Systematic Screening for Behavior Disorders (SSBD): A multiple gating procedure: Technical manual*. Longmont, CO: Sopris West.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (SSBD)* (2nd ed.). Longmont, CO: Sopris West.

Wehmeier, P. M., Schacht, A., & Barkley, R. A. (2010). Social and emotional impairment in children and adolescents with ADHD and the impact on quality of life. *The Journal of Adolescent Health, 46*, 209–217. http://dx.doi.org/10.1016/j.jadohealth.2009.09.009

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*, 806–838. http://dx.doi.org/10.1177/0011000006288127

Received June 30, 2014
Revision received October 16, 2014
Accepted October 18, 2014 ■