

An Extension Convergent Validity Study of the Systematic Screening for Behavior Disorders and the Achenbach Teacher's Report Form With Middle and High School Students With Emotional Disturbances

Gregory J. Benner, Brad M. Uhing, Corey D. Pierce, Kathleen M. Beaudoin,
Nicole C. Ralston, and Paul Mooney

Abstract: We sought to extend instrument validation research for the Systematic Screening for Behavior Disorders (SSBD) (Walker & Severson, 1990) using convergent validation techniques. Associations between Critical Events, Adaptive Behavior, and Maladaptive Behavior indices of the SSBD were examined in relation to syndrome, broadband, and total scores of the Achenbach Child Behavior Checklist-Teacher's Report Form (TRF) (Achenbach, 2001). Both measures were conducted with 65 students with emotional disturbance in grades 6 through 12. Overall convergent validity of the SSBD and TRF was strong, particularly for TRF externalizing problems and associated syndromes. Results provide further support for use of the SSBD in the assessment of behavioral functioning of students with emotional disturbance and extend validation for use of this instrument with secondary students.

Introduction

During the past 10 years, there has been a 20% increase in the number of children identified with emotional disturbance (ED) under the Individuals with Disabilities Education Act (IDEA) (U.S. Department of Education, 2002). U.S. public schools provide special education and related services to nearly 500,000 students labeled with emotional disturbance (U.S. Department of Education, 2002). Although 52% of students with disabilities graduated with a regular high school diploma in 2003, only 35% of students with ED did so. Furthermore, 56% of students with ED dropped out of school in 2003, substantiating the claim that students with ED consistently have the lowest graduation rates and highest dropout rates of students in the public school system (U.S. Department of Education, 2005). Consequently, students with ED continue to face problems throughout their teenage and adult years, including enhanced risk for arrest and substance abuse, job instability, higher usage of welfare and mental health services, and limited income earnings (Mayer, Lochman, & Van Acker, 2005; Rock, Fessler, & Church, 1997).

Because of their low rates of success in public schools and bleak long-term outcomes, it is apparent that students with ED present a variety of complex and challenging behaviors (Cullinan, 2007). For example, the current definition of ED in IDEA interprets the term emotional disturbance to mean one or more of a series of five "characteristics" that

are present "over a long period of time and to a marked degree" and "adversely affect a student's educational performance" (U.S. Department of Education, 2002). These characteristics include the following: an inability to learn that cannot be explained by intellectual, sensory, or health factors; an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; inappropriate types of behavior or feelings under normal circumstances; a general pervasive mood of depression; and a tendency to develop physical symptoms or fears associated with personal or school problems. School multidisciplinary teams are faced with the challenge of designing treatment programs that meet the behavioral and academic needs of students with ED. As a result, it is critical that decisions made on behalf of students with ED are based on accurate assessment data.

It can be challenging to determine whether a student fits the IDEA definition of ED (Cullinan, Osborne, & Epstein, 2004). Thus, it is important that steps be taken to validate formal instruments used in the assessment of ED. Instruments used in the assessment of ED should be highly reliable and valid so that useful data are gathered for decision-making purposes. For example, assessment instruments should be able to provide a holistic view of student's social-emotional functioning for planning and implementing effective treatment interventions.

In school-based assessment, behavior rating scales are one of the primary methods used to

identify students with ED (Mattison, 2001). Behavior rating scales have become extremely popular because of their ease of administration, time, and cost efficiency, and ability to monitor the current status and functioning of students with ED as well as to monitor their outcomes over time. Additionally, the use of rating scales in assessment allows for multiple informants (i.e., parents, teachers, students) to assess the functioning of students, which typically provides a broader range of perspectives on that student's behavior (Achenbach & McConaughy, 1996; Mash & Wolfe, 1999).

One of the most widely used rating scales for assessing social-emotional functioning is the *Child Behavior Checklist-Teacher's Report Form* (TRF) (Achenbach, 2001). The TRF is a standardized, norm-referenced behavior rating scale for teachers which assesses the social adjustment of students. The TRF is primarily a problem checklist consisting of 113 items. Teachers are asked to rate students on a variety of behaviors, and the instrument provides two broadband scores, "internalizing" and "externalizing," plus a "total scale" score for each participant. The TRF also provides score profiles on eight syndromes: Aggressive Behavior, Anxious/Depressed Behavior, Attention Problems, Delinquent Behavior, Social Problems, Somatic Complaints, Thought Problems, and Withdrawn Behavior. Students who score in the borderline clinical range or higher on one or more of the syndromes or on the overall index are considered at risk for behavioral difficulties.

Validity refers to a test's ability to measure what it purports to measure (Salvia & Ysseldyke, 2004). Valid instruments are critical in assessing students for ED and, if used appropriately for their intended purposes, assist practitioners in gathering data that allows for confidence in the decision-making process (Sattler, 2001). A number of different methods of examining the validity of an instrument are appropriate. One of the methods of examining validity is called convergent validity. Convergent validity examines the relationship between assessment instruments that measure the same constructs (Salvia & Ysseldyke, 2004). Demonstrating the convergent validity of an assessment instrument can increase the confidence that results obtained from that instrument reflect the constructs intended to be measured by that instrument. Thus, the higher the relationship between the two instruments, the stronger the convergent validity (Epstein, Nordness, Nelson, & Hertzog, 2002).

Existing convergent validity data provide support for the use of the TRF in assessing students' social-emotional functioning. Harniss and colleagues (Harniss, Epstein, Ryser, & Pearson, 1999) examined the convergent validity of the TRF and the *Behavioral and Emotional Rating Scale* (BERS) (Epstein & Sharma, 1998) in adolescents with ED. Specifically, the five positively based subscales and overall strength index of the BERS were correlated to the competence scales, internalizing and externalizing broadband dimensions, and total problem score of the TRF. Correlations ranged from moderately (.39) to highly (.72) positive for the competence scales. Correlations were moderately to highly negative with the externalizing broadband dimension and generally low for the internalizing broadband dimension. Meanwhile, Trout, Ryan, La Vigne, & Epstein (2003) sought to replicate the Harniss et al. (1999) study on an early childhood sample of students. Again, correlations were moderately to highly positive across the BERS and TRF subscales, ranging from .29 to .73. Additionally, the BERS evidenced moderately to highly negative correlations when compared to the

TRF internalizing and externalizing broadband dimensions, ranging from -.23 to -.62. In a third study (Emerson, Crowley, & Merrell, 1994), the convergent validity of the TRF and *School Social Behavior Scales* (SSBS) (Merrell, 1993) was examined on fourth- and fifth-grade public school students who were primarily Caucasian (95%). Specifically, the adaptive functioning subscale and the internalizing and externalizing broadband dimensions of the TRF were correlated with the social competence (Scale A) and Antisocial Behavior (Scale B) scales of the SSBS. As expected, correlations were moderate to high and in the expected directions when comparing the SSBS social competence subscales with the TRF adaptive functioning subscale (.65 to .73), internalizing broadband dimension (-.57 to -.62), and externalizing broadband dimension (-.55 to -.75). Additionally, correlations were also moderate to high and in the expected directions when comparing the SSBS antisocial behavior subscales with the TRF adaptive functioning scale (-.45 to -.62), internalizing broadband dimension (.34 to .52), and externalizing broadband dimension (.76 to .84).

The *Systematic Screening for Behavior Disorders* (SSBD) (Walker & Severson, 1990) is a three-stage screening process that was originally designed for the screening of social and emotional behavioral problems of elementary school students. Stage I includes teacher nominations and rank-ordering of pupils meeting specific definitions of behavior difficulties; Stage II includes teacher completion of the Adaptive and Maladaptive Behavior rating scales; and Stage III includes observation of the student in various settings. The SSBD has demonstrated mixed results with respect to the technical adequacy of the instrument. Zlompke and Spies (1998) reviewed the SSBD and found several studies presented in the manual that support the development and validation of the SSBD, although a few correlations were less than desirable. Stage I test-retest rank order correlations (one-month retest) averaged .76 for externalizers and .74 for internalizers, respectively. However, of the top three students listed for externalizing and internalizing behaviors, only 69% were listed among the top three students a month later. During Stage II trial testing, test-retest reliabilities were much higher and improved to .88 for adaptive and .83 for maladaptive behaviors (Zlompke & Spies, 1998). Similar results were found within measures of internal consistency for Stage II, with coefficient alphas averaging .86 on the adaptive and .84 maladaptive scales, respectively. Analyses were not conducted on the Stage III due to low frequencies of positively checked items (Zlompke & Spies, 1998). However, the researchers reported that interrater agreement ratios for Stage III were consistently within the .80 to .90 range (using 10-second interval recording).

Discriminant validity studies of the SSBD support the use of the instrument in areas such as classifying group membership (e.g., ED versus non-ED populations) and discriminating between students' externalizing and internalizing behaviors (Zlompke & Spies, 1998). However, predictive and concurrent validity studies suggest the instrument has low to moderate correlations in Stage I, II, and III measures. For instance, predictive validity data indicated that on Stage I measures, only 52% of internalizers and 69% of externalizers from the previous year were listed among the top three ranked students in the following year (Zlompke & Spies, 1998). Stage II correlations ranged from .32 (Critical Events Index) to .70 (Maladaptive Rating Scale), respectively, and when shared with the Stage III measure, indicated classification efficiencies in the low to moderate range (Zlompke & Spies, 1998).

Concurrent validity data was addressed in the manual by correlating the total score on the Stage II ratings with other measures designed by the first author (Walker-McConnell Scale of Social Competence and School Adjustment). While these data suggest there is some support for the Stage II measures of the SSBD, most scores were also in the low to moderate range (Zlompke & Spies, 1998).

Recently, researchers have extended the use of the SSBD to middle and junior high school students with positive results (Caldarella, Young, Richardson, Young, & Young, 2008; H. M. Walker, personal communication, June 21, 2007). For example, Caldarella, Young, Richardson, Young, and Young (2006, 2008) asked teachers of students in grades six through nine to identify students at risk for emotional and behavioral difficulties (SSBD Stage 1). Teachers completed SSBD Stage 2 scales (Critical Events Index, Maladaptive Behavior, and Adaptive Behavior) as well as the TRF and the Social Skills Rating Scale (Gresham & Elliot, 1990) on 123 students meeting teacher nomination criteria at SSBD Stage 1. Caldarella and colleagues (2008) found small to moderate correlations between TRF Externalizing and Internalizing and SSBD scales. Correlations between SSBD Adaptive Behavior and TRF Externalizing and Internalizing Scales were small in magnitude ($-.33$ and $.17$, respectively); whereas those between SSBD Maladaptive Behavior and TRF Externalizing and Internalizing Scales were moderate and small in magnitude ($.67$ and $-.23$, respectively). Using t-test comparisons of each item between students nominated as internalizing or externalizing on SSBD Stage 1, Caldarella and colleagues (2008) divided items on the Critical Events Index (CEI) into either internalizing or externalizing categories. Correlations between SSBD CEI Externalizing and TRF Externalizing and Internalizing Scales were moderate in magnitude ($.51$ and $-.30$, respectively). Similarly, moderate correlations between SSBD CEI Internalizing and TRF Externalizing and Internalizing Scale were found ($-.37$ and $.53$, respectively). Findings indicated that the SSBD shows promise as a valid and reliable screening measure for at-risk secondary students (Caldarella et al., 2006, 2008).

Current studies suggest that the TRF compares favorably with other measures, including the BERS and SSBS. However, previous studies have compared the TRF to other measures using limited populations, most notably, adolescent students, early childhood students, and non-disabled fourth- and fifth-grade Caucasian students. To date, researchers have not examined the convergent validity of the widely used behavioral screening measure *Systematic Screening for Behavior Disorders* (Walker & Severson, 1990) with other standardized measures of behavioral functioning on secondary populations with ED. The primary purpose of this study was to examine the convergent validity of the SSBD with the TRF on a sample of sixth through twelfth-grade public school students receiving special education services for ED served in self-contained settings.

Method

Participants

Sixty-five public school students (51 males and 14 females) receiving special education services for ED in an urban, northwestern city participated in this study. The participants were served across nine different settings: one middle school ($n = 8$), three high schools ($n = 46$), one psychiatric residential treatment facility ($n = 5$), and one

interim alternative educational setting ($n = 6$). Ethnic breakdowns were 43% Caucasian ($n = 28$), 25% African-American ($n = 16$), 9% Hispanic ($n = 6$), 5% Native-American ($n = 3$), 1% Asian ($n = 1$), and 17% mixed ethnicity ($n = 11$). The specific number and approximate percentage of the 65 participants at each grade level follows: sixth grade, $n = 3$ (5%); seventh grade, $n = 6$ (9%); eighth grade, $n = 3$ (5%); ninth grade, $n = 21$ (32%); tenth grade, $n = 19$ (29%); eleventh grade, $n = 10$ (15%); and twelfth grade, $n = 3$ (5%). Ages of students ranged from 12 to 20 years, with a mean of 16.0 ($SD = 1.8$).

Thirteen teachers of participating students completed ratings of students' social and emotional strengths and problem behaviors. The number of teachers employed at the middle and high school grade levels were two (22%) and seven (78%), respectively. Six teachers were female (67%) and three were male (33%). The number of years teaching students with ED ranged from 2 to 28, with an average of 10.5 years ($SD = 10.2$). All participating teachers held special education teaching endorsements. Teacher caseloads ranged from 8 to 21 students, with a mean of 11.9 ($SD = 5.1$).

Measures

The *Systematic Screening for Behavior Disorders* (SSBD) (Walker & Severson, 1990; 1992) is a three-stage screening process that begins with teacher nominations and rank-ordering of pupils meeting specific definitions of behavior difficulties. The second stage consists of a 33-item Critical Events Index (CEI) checklist and a 23-item Combined Frequency Index (CFI) checklist. The CEI contains 33 items measuring low-frequency, high-intensity behavior problems (e.g., sets fires, steals). The respondent indicates whether the critical event has or has not occurred within the past six months. The CFI consists of two behavior-rating scales: Adaptive Behavior (12 items) and Maladaptive Behavior (11 items). The Adaptive Behavior scale includes 12 items that assess classroom and peer adaptive adjustment (e.g., is considerate of the feelings of others). The Maladaptive Behavior scale has 11 items that focus on school-related behavior problems (e.g., refuses to participate in games and activities with other children at recess). Both the Adaptive and Maladaptive scales measure the frequency of the student's behavior within the past month.

The *Child Behavior Checklist-Teacher's Report Form* (Achenbach, 2001) is used to measure the social adjustment of participants. The TRF consists of 113 problem items, such as difficulty following directions, disturbs other pupils, and disrupts class discipline. The teacher rates the child on each item and indicates the severity of the problem on a three-point Likert-type scale ranging from 0 (Not True) to 2 (Very True or Often True). The TRF scoring profile provides a total scale score (Total Problems), two broadband scale scores (Internalizing and Externalizing), and eight syndrome subscale scores (Withdrawn Behavior, Somatic Complaints, Anxious/Depressed Behavior, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior). The broadband Internalizing scale score is based on the sum of the Withdrawn Behavior, Somatic Complaints, and Anxious/Depressed Behavior scale scores. The broadband Externalizing scale score is based on the Rule-Breaking Behavior and Aggressive Behavior scale scores. The Social Problems, Thought Problems, and Attention Problems syndrome subscale scores are

not included on either the broadband Internalizing or Externalizing scale scores. The TRF test-retest and internal consistency values for the broad and syndrome scales are reported in the test manual as ranging from .62 to .96 and .72 to .95, respectively (Achenbach, 1991). The syndrome and broadband scale scores of participants in the present study indicated very strong internal consistency with a Cronbach's Alpha of .95.

Procedures

Thirteen special education teachers serving students with ED in self-contained classrooms completed the SSBD and TRF for each participating student in May of 2005. Teachers did not complete SSBD or TRF protocols for students whom they had known for less than two months. A two-hour training session familiarized teachers with the structure (i.e., item formats) and specific instructions for completing these measures. Teachers were given two weeks to complete the two scales. Each student was rated independently by teachers. Research assistants were trained to score and enter the data derived from SSBD and TRF protocols. The training and scoring reliability procedures used with research assistants follow.

Training. Two research assistants completed the scoring of SSBD and TRF protocols. The research assistants reached 100% fidelity in scoring each measure on three consecutive trials. Scoring fidelity was determined by comparing the research assistants' scoring of a practice protocol with one scored accurately. When the research assistants reached the fidelity criterion they began scoring the SSBD and TRF protocols of participating students.

Scoring reliability. Scoring reliability checks on all SSBD and TRF protocols were conducted at two phases of data collection. First, each protocol was checked for scoring accuracy by two of the authors after initial scoring by research assistants. More specifically, each protocol was checked to determine that items were completed, raw scores were computed accurately for each subtest, and standard scores were derived accurately. Agreement was calculated by dividing the number of agreements by agreements plus disagreements and multiplying by 100. An agreement was recorded when the agreement check calculations aligned with calculations made after initial scoring. Agreement in scoring SSBD and TRF protocols was 98% (range = 96% to 100%), and 99% (range = 98% to 100%), respectively. Second, all scores were checked for accuracy by researchers following initial data entry. Agreement in entering SSBD and TRF data was 99%. Initial errors made in scoring or data entry were corrected.

Results

This study utilized Pearson's Product-Moment correlation coefficients to analyze the relationship between the SSBD scale and the TRF syndrome, broadband, and total scores of the 65 participating youth (see Table 1). The 11 correlations between the SSBD Adaptive Behavior scale and TRF syndrome, broadband, and total scores were negative, whereas all of the remaining 22 correlations were positive and in the expected direction. The strength of the correlations varied from .29 to .89. The strength, or magnitude, of correlations was assessed using the scale developed by Hopkins (2002). Correlations of .1 to .29, .3 to .49, .5 to .69, .70 to .89, and .90 or more were considered small, moderate, large, very large, and nearly perfect,

Table 1

Correlations Between SSBD Scales and TRF Syndrome and Composite Scores

	SSBD Scale		
	Critical Events	Adaptive Behavior	Maladaptive Behavior
TRF Syndrome Scores			
Anxious/Depressed	.63**	-.35**	.35**
Withdrawn	.61**	-.44**	.29*
Somatic Complaints	.76**	-.58**	.52**
Social Problems	.69**	-.54**	.61**
Thought Problems	.80**	-.62**	.50**
Attention Problems	.79**	-.81**	.76**
Rule-Breaking Behavior	.53**	-.78**	.73**
Aggressive Behavior	.75**	-.78**	.89**
TRF Broadband Scores			
Internalizing	.75**	-.53**	.48**
Externalizing	.72**	-.84**	.89**
TRF Total Score			
Total Problems	.82**	-.83**	.79**

Note. * $p < .05$ and ** $p < .01$.

respectively. Using these criteria, 17 correlations (52%) were very large, 11 (33%) were large, 4 (12%) moderate, and 1 (3%) small in magnitude. Thus, with one exception, correlations were moderate to very large in magnitude.

Several correlations warrant highlighting. Very large (i.e., .70 to .89) correlations were found between the TRF Total Problems and the SSBD Critical Events Index ($r = .82, p < .01$), Adaptive Behavior scale ($r = -.83, p < .01$), and Maladaptive Behavior scale ($r = .79, p < .01$). Very large correlations were also found between the TRF Externalizing Problems and the SSBD Critical Events Index ($r = .72, p < .01$), Adaptive Behavior Scale ($r = -.84, p < .01$), and Maladaptive Behavior Scale ($r = .89, p < .01$). The correlation between the TRF Internalizing Problems and SSBD Critical Events Index was also very large ($r = .75, p < .01$). The magnitude of the correlations between the TRF Internalizing Problems and the SSBD Adaptive Behavior ($r = -.53, p < .01$), and Maladaptive Behavior scales ($r = .51, p < .01$) were large and moderate, respectively. This indicates that the overall convergent validity of the SSBD and the TRF was very strong, par-

ticularly for TRF Externalizing Problems and associated syndromes. For example, very large correlations were found between the TRF Aggressive Behavior externalizing syndrome and the SSBD Critical Events Index ($r = .75, p < .01$), Adaptive Behavior Scale ($r = -.78, p < .01$), and Maladaptive Behavior Scale ($r = .89, p < .01$).

Another framework for determining validity is provided by Anastasi and Urbina (1996). These researchers report that in order for a correlation coefficient to be cited as evidence of validity, it should demonstrate statistical significance. As indicated in Table 1, all of the 33 correlations meet this criterion (i.e., $p < .05$).

Discussion

Continuing to research the validity of behavior rating scales such as the SSBD serves a function in the field of assessment of behaviors. Schools are conscious of the difficulty in accurately identifying the presence of behavior disorders in students (Cullinan, et al., 2004; Uhing, Mooney, & Ryser, 2005) and rely heavily upon behavior rating scales to identify students who would benefit from behavioral supports to improve school performance (Mattison, 2001). The purpose of this study was to extend the validation evidence for the SSBD by examining the convergent validity of the SSBD and the TRF with students in 6th through 12th grade with ED.

Previous convergent validity studies conducted between the TRF and other behavior rating scales resulted in largely moderate to high correlations (Harniss et al., 1999; Trout et al., 2003). Results of the present study reaffirmed those previous studies, demonstrating a range of correlations from small ($r = .29$) to very large ($r = .89$) in magnitude. In Harniss et al. (1999) and Trout et al. (2003), the patterns indicated that stronger correlations were generally reported for TRF externalizing versus internalizing domains. Results also extended evidence of validity for the use of the TRF with the full range of school-age students. Whereas previous convergent validity studies demonstrated evidence for specific, limited age groups of students, none of whom were labeled as having an IDEA disability, the current evidence was gained using a population of students that ranged from sixth to twelfth grade and included students identified with an IDEA disability. Demonstrating moderate to high correlations is an important component in determining the validity of an instrument in a convergent validity study. It is also critical to document the statistical significance of those correlations. The results showed that 100% ($N = 33$) of the correlations were significant at the $p < .01$ level.

Our findings extend the validation of the SSBD to middle and high school students with ED. In their sample of 123 middle and junior high school students, Caldarella and colleagues (2008) found small to moderate correlations between TRF Externalizing and Internalizing and SSBD Critical Events, Maladaptive, and Adaptive scales. We extend the work of Caldarella and colleagues (2008) by sampling from middle and high school students receiving services for ED, placed in self-contained settings. We found that, with one exception, the 33 correlations between the TRF and SSBD scales were moderate to very large in magnitude. Caldarella and colleagues (2008) found that correlations between SSBD Adaptive Behavior and TRF Externalizing and Internalizing Scales were small in magnitude ($-.33$ and $.17$, respectively); whereas we found very large ($-.84$) and large ($.53$) correlations, respectively. In addition, Caldarella and colleagues (2008)

found moderate and small correlations between SSBD Maladaptive Behavior and TRF Externalizing and Internalizing Scales ($.67$ and $-.23$, respectively); whereas we found very large ($.89$) and moderate correlations ($.48$), respectively. It remains unclear what variables might explain the more robust correlations between the SSBD and TRF found in the present study. One explanation may be the nature of participants in each study. Students in the present investigation met inclusion criteria by being formally identified with emotional disturbance and in being served in a self-contained placement whereas those participating in Caldarella et al. (2008) met SSBD criteria for risk of behavioral disorder.

Four primary limitations within this study should be acknowledged. First, the nine public schools in which all of the participants were enrolled were located within the same northwestern city in the United States. It is recognized that had the study included a more diverse population of schools from across the U.S., the results of this study may have been different. Second, the largest percentage of participants included in this study was from the middle school grade level (i.e., grades 7, 8, and 9). It is plausible that the results may be different if the same study was repeated with a sample more evenly distributed across grade levels. Third, it is possible that these results may not generalize beyond students with ED, as all participants were identified as having an ED diagnosis. Finally, no observations were conducted and no permanent products were collected to assess validity in this study.

These limitations can be addressed in two ways. First, professionals interested in the continued effort to validate assessment instruments could collaborate in an effort to collect data from a more diverse population of students, including those from different geographic locations. Second, future convergent validity research in the area of behavior could include more direct measures of behavior.

There are two primary implications of the present study. First, the role of validating assessment instruments is shared by both researchers and practitioners (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The findings of the present study supporting the validity of the SSBD and the TRF not only add to the evidence base for these instruments but also support the notion that practitioners engaged in ongoing professional development and interventions can contribute to the literature in this area of study. It demonstrates that practitioners active in the field can and may become more engaged in the research process in an effort to continue to determine the soundness of the instruments they are using in the field. Second, although universal screening for basic reading, mathematics, and writing skills is relatively straightforward and efficient because there are well-established measures and benchmark standards for performance available to schools, this is not the case for social behavior. The SSBD is the only available universal screening instrument for social behavior. Our findings extend the extant validation literature of the SSBD with secondary students formally identified with ED. Our findings suggest the potential use of the SSBD as a valid measure of the behavioral functioning of students with emotional disturbance from primary through secondary grade levels. These data may inform the implementation of Response to Intervention (RtI) in the area of social behavior, particularly among students with the most challenging behaviors.

References

- Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (2001). *Teacher's Report Form for ages 6-18*. Burlington, VT: University of Vermont.
- Achenbach, T. M., & McConaughy, S. H. (1996). Relations between DSM-IV and empirically based assessment. *School Psychology Review, 25*(3), 329-341.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Macmillan.
- Caldarella, P., Young, E. L., Richardson, M. J., Young, B. J., & Young, K. R. (2006, March). *School-wide screening for students with behavioral and emotional concerns*. Poster presented at the Third International Conference on Positive Behavior Support, Reno, NV.
- Caldarella, P., Young, E. L., Richardson, M. J., Young, B. J., & Young, K. R. (2008). Validation of the systematic screening for behavior disorders in middle and junior high school. *Journal of Emotional and Behavioral Disorders, 16*(2), 105-117.
- Cullinan, D. (2007). *Students with emotional and behavioral disorders: An introduction for teachers and other helping professionals* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Cullinan, D., Osborne, S., & Epstein, M. H. (2004). Characteristics of emotional disturbance among female students. *Remedial and Special Education, 25*(5), 276-290.
- Emerson, E. N., Crowley, S. L., & Merrell, K. W. (1994). Convergent validity of the School Social Behavior Scales with the Child Behavior Checklist and Teacher Report Form. *Journal of Psychoeducational Assessment, 12*, 372-380.
- Epstein, M. H. (2004). *Behavioral and Emotional Rating Scale*, (2nd ed.). Austin, TX: PRO-ED.
- Epstein, M. H., Nordness, P. D., Nelson, J. R., & Hertzog, M. (2002). Convergent validity of the Behavioral and Emotional Rating Scale with primary grade-level students. *Topics in Early Childhood Education, 22*(2), 114-121.
- Epstein, M. H., & Sharma, J. (1998). *Behavioral and Emotional Rating Scale: A strength-based approach to assessment*. Austin, TX: PRO-ED.
- Gresham, F. M., & Elliott, S.N. (1990). *Social Skills Rating System Manual*. Circle Pines: American Guidance Service.
- Harniss, M. K., Epstein, M. H., Ryser, G., & Pearson, N. (1999). The Behavioral and Emotional Rating Scale: Convergent validity. *Journal of Psychoeducational Assessment, 17*, 4-14.
- Hopkins, W. (2002). A new view of statistics. Retrieved March 29, 2007 from <http://www.sportsci.org/resource/stats/effectmag.html>.
- Mash, E., & Wolfe, D. (1999). *Abnormal child psychology*. Belmont, CA: Wadsworth Publishing.
- Mattison, R. (2001). Longitudinal use of the teacher's report form in tracking outcomes for students with SED. *Journal of Emotional and Behavioral Disorders, 9*(2), 86-93.
- Mayer, M., Lochman, J., & Van Acker, R. (2005). Introduction to the special issue: Cognitive-behavioral interventions with students with EBD. *Behavioral Disorders, 30*(3), 197-212.
- Merrell, K. W. (1993). *School social behavior scales*. Brandon, VT: Clinical Psychology.
- Rock, E. E., Fessler, M. A., & Church, R. P. (1997). The concomitance of learning disabilities and emotional/behavioral disorders: A conceptual model. *Journal of Learning Disabilities, 30*(3), 245-263.
- Salvia, J., & Ysseldyke, J. (2004). *Assessment in special and inclusive education* (9th ed.). Boston: Houghton and Mifflin.
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Jerome M. Sattler, Publisher.
- Trout, A. L., Ryan, J. B., La Vigne, S. P., & Epstein, M. H. (2003). Behavioral and Emotional Rating Scale: Two studies of convergent validity. *Journal of Child and Family Studies, 12*(4), 399-410.
- Uhing, B. M., Mooney, P., & Ryser, G. R. (2005). Differences in strength assessment scores for youth with and without ED across the youth and parent ratings scales of the BERS-2. *Journal of Emotional and Behavioral Disorders, 13*, 181-187.
- U.S. Department of Education (2002). *Twenty-fourth annual report to Congress on the implementation of the education of the handicapped act*. Washington, DC: U.S. Government Printing.
- U.S. Department of Education. (2005). *Twenty-seventh annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- Walker, H. M., & Severson, H. H. (1990). *Systematic screening for behavior disorders* (2nd ed.). Longmont, CO: Sopris West.
- Walker, H. M., & Severson, H. H. (1992). *Technical manual, systematic screening for behavior disorders*. Longmont, CO: Sopris West.
- Zlompeke, L. C., & Spies, R. (1998). Test review of the systematic screening for behavior disorders. From B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The thirteenth mental measurements yearbook* [electronic version]. Retrieved December 9, 2007, from <http://www.unl.edu/buros>.
-

Authors

Gregory J. Benner, Ph.D., is an associate professor of Special Education at the University of Washington, Tacoma. His current research interests include preventive systems change to meet student academic and social/emotional needs, literacy and language instruction for diverse learners, and building the academic achievement of students with or at risk of emotional disturbance.

Brad M. Uhing, Ph.D., is a Research Scientist at the Avera Research Institute in Sioux Falls, South Dakota. His research interests include early intervention programs for children at-risk for emotional disturbance, assessment of emotional disturbance, and family genetic factors related to emotional disturbance.

Corey D. Pierce, Ph.D., is an Assistant Professor in the Special Education Program at University of Northern Colorado. His current research interests include the academic performance of students with behavior disorders and learning difficulties, whole school models for delivering a continuum of academic opportunities to meet the diverse needs of all learners, and the assessment of students for special education eligibility.

Kathleen M. Beaudoin, Ph.D., is an Associate Professor in the Education Program at University of Washington, Tacoma. Her areas of expertise include training pre-service and practicing educators to work with students with emotional and behavioral disorders.

Nicole C. Ralston, M.Ed., is a research analyst and doctoral candidate in the Education Department at the University of Washington.

Paul Mooney, Ph.D., is an Assistant Professor in the Department of Educational Theory, Policy, and Practice at Louisiana State University. His research interests include academic assessments and interventions for students with or at-risk for emotional and behavioral disorders.
