

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Special Education and Communication Disorders  
Faculty Publications

Department of Special Education and  
Communication Disorders

---

2015

# Validation of the Symptoms and Functioning Severity Scale in Residential Group Care

Matthew C. Lambert

*University of Nebraska-Lincoln*, mlambert2@unl.edu

Kristin Duppong-Hurley

*University of Nebraska-Lincoln*, kristin.hurley@unl.edu

Thomas J. Gross

*University of Nebraska-Lincoln*, tjgross44@gmail.com

Michael Epstein

*University of Nebraska - Lincoln*, mepstein1@unl.edu

Amy L. Stevens

*Boys Town 100 Crawford Drive Boys Town*, amy.stevens@boystown.org

Follow this and additional works at: <http://digitalcommons.unl.edu/specedfacpub>



Part of the [Special Education and Teaching Commons](#)

---

Lambert, Matthew C.; Duppong-Hurley, Kristin; Gross, Thomas J.; Epstein, Michael; and Stevens, Amy L., "Validation of the Symptoms and Functioning Severity Scale in Residential Group Care" (2015). *Special Education and Communication Disorders Faculty Publications*. 119.

<http://digitalcommons.unl.edu/specedfacpub/119>

This Article is brought to you for free and open access by the Department of Special Education and Communication Disorders at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Special Education and Communication Disorders Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Published in final edited form as:

*Adm Policy Ment Health*. 2015 May ; 42(3): 356–362. doi:10.1007/s10488-014-0575-z.  
© Springer Science+Business Media New York 2014. Used by permission.

## Validation of the Symptoms and Functioning Severity Scale in Residential Group Care

**Matthew C. Lambert,**

University of Nebraska – Lincoln 273 Barkley Memorial Center Lincoln, NE 68583  
mlambert2@unl.edu 402-472-5487

**Kristin Duppong Hurley,**

University of Nebraska – Lincoln 247E Barkley Memorial Center Lincoln, NE 68583  
kristin.hurley@unl.edu 402-472-5501

**Thomas J. Gross,**

University of Nebraska – Lincoln 213 Barkley Memorial Center Lincoln, NE 68583  
tjgross44@gmail.com 402-472-5486

**Michael H. Epstein, and**

University of Nebraska – Lincoln 202F Barkley Memorial Center Lincoln, NE 68583 402-472-5472

**Amy L. Stevens**

Boys Town 100 Crawford Drive Boys Town, NE 68010 amy.stevens@boystown.org

### Abstract

Tests that measure the emotional and behavioral problems of children and youth are typically not normed and standardized on youth diagnosed with disruptive behavior, particularly those youth in residential care. Yet professional standards mandate that before instruments are used with a specific population the psychometric properties need to be studied and re-established: specifically, psychometric properties, including validity, need to be evaluated (AERA, APA, & NCME, 1999). The purpose of the present study was to assess the validity characteristics of the Symptoms and Functioning Severity Scale (SFSS; Bickman, et al., 2010), a widely used test developed for use in outpatient clinics, with youth in a residential care program. The convergent validity of the SFSS was established with the large correlations (.78-.86) with the CBCL. Several binary classification analyses including specificity, area under the receiver operating characteristic curve, positive and negative likelihood ratios, and the Youden Index supported the validity of the SFSS. However, the sensitivity index was somewhat low indicating the test may produce a high level of false negatives. Limitations, future research and implications are discussed.

---

Youths at-risk for externalizing behavioral problems are of special concern to researchers due to the strain placed upon their family units, the poor outcomes they tend to experience, and their large consumption of mental health resources such as out-of-home placements (Brown et al., 2011; Webster-Stratton, 1993). These behaviors are frequently described as physical and verbal aggression, noncompliance, disregard for rules or others rights, impulsivity, and excessive activity (Kolko & Pardini, 2010). Youth with these problems often meet diagnostic criteria for conduct disorder (CD), oppositional defiant disorder (ODD), and/or attention deficit hyperactivity disorder (ADHD; McKinny & Morse, 2012;

Trentacosta, Hyde, Goodlett, & Shaw, 2013). Additionally, a large proportion of these youth may require out-of-home placement, such as residential care or treatment foster care (Brown et al., 2011; Cuthbert et al., 2011). Externalizing problems of youth tend to persist and increase over time without identification and treatment (Short & Shapiro, 1993; Webster-Stratton, 1993) and frequently result in educational, vocational, and financial failure (Charles, Bywater, & Edwards, 2011; Bradley, Doolittle, & Bartolotta, 2008).

The impact of externalizing behavioral problems highlights the importance of assessments that quickly and reliably identify children at-risk for these problems. Moreover, if rating scales are used they should include multiple perspectives so typical behavioral and emotional functioning may be reliably assessed (McKinny & Morse, 2012). Parent, caregiver, teacher, and clinician reports can provide indicators of baseline functioning at home and in community settings, facilitate comparisons to peers, and function as a progress monitoring tool during care. Youth reports add information about the subjective experience of children and provide diagnostic clarity (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005).

## Assessment and Progress Monitoring

There are multiple norm-referenced measures for evaluating youth emotional and behavioral problems or functioning. Some well known measures include the 118-item Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), the 100 to 160-item Behavior Assessment System for Children – Second Edition (BASC-2; Reynolds & Kamphaus, 2004), and the 52-item Behavioral and Emotional Rating Scale-2 (BERS-2; Epstein, 2004). These measures include parent, teacher, and child-report forms, and may be used for progress monitoring. Unfortunately, these measures may not be suited for frequent behavioral problem or functioning assessment because they may become cumbersome to repeatedly complete. Brief measures of behavioral and emotional symptoms may be more feasible for routine assessments.

The 25-item Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is one example of a brief multi-rater assessment tool, which was designed to evaluate child behavioral and emotional functioning (Goodman, 1997). The authors of the measure express that the SDQ is meant to screen for potential problems or the likelihood of psychiatric diagnoses (Goodman, 1997). Another brief measure of behavioral and emotional functioning is the Symptoms and Functioning Severity Scale (SFSS; Bickman et al., 2010). The SFSS is a 24-item youth behavior rating scale that may be completed by a clinician, adult caregiver, or youth. The items are divided into two subscales, Externalizing Problems (14 items) and Internalizing Problems (10 items), and yields a Total Problems score. The SFSS was developed specifically to measure changes in emotional and behavioral symptoms during the course of treatment (Bickman et al., 2010). The SFSS has advantages that would make it useful for routine progress monitoring: (a) it is brief; (b) it is sensitive to change (Athay, Riemer & Bickman, 2012); (c) it provides data for clinical decision-making; and (d) it can be used with different raters and across a variety of treatment settings (Athay et al., 2012; Duppong Hurley, Lambert, & Stevens, 2013).

The SFSS would contribute to assessment and progress monitoring by providing a brief yet comprehensive assessment of behavioral problems and functioning for youths during their course of treatment. Caregivers may provide important ratings of youth behaviors at home and in the community. Clinicians may give information regarding how others perceive youth behaviors. Youth may give information regarding their subjective experience of behavioral problem severity. Still, the first step in establishing its use is demonstrating if it is psychometrically sound with various youth populations. SFSS internal consistency estimates with an outpatient sample were found to be good to excellent for the Internalizing subscale, Externalizing subscale, and Total Problems scale ( $\alpha = .89, .93, \text{ and } .92$ , respectively; Athay et al., 2012). The factor structure and internal consistency of the SFSS was replicated with a sample of youth receiving residential group care with externalizing behavioral problems (Duppong Hurley et al., 2013). The correlations between the SFSS parent and youth forms, and the ASEBA's CBCL and Youth Self Report had similarly robust correlations ( $r = .86$  and  $.77$ , respectively; Athay et al., 2012). However, the validity of the SFSS has not been replicated with high-need youth populations.

Based on prior research from Athay et al. (2012) and Duppong Hurley et al. (2014), scores from the SFSS appear to be reliable for assessing youth behavioral problems including those youth in residential group care. However, more research is needed to confirm the validity of both the continuous and categorical scores from the SFSS. Further, the classification properties of the SFSS have not been examined. The purpose of this study was two-fold: (1) to evaluate the degree to which the SFSS continuous and categorical scores converge with scores from the CBCL for youth with high needs, and (2) to examine the ability of the SFSS to classify youth with regard to emotional and behavioral needs. Additional support for the validity and classification properties of the SFSS would allow service providers the option of a brief, yet psychometrically sound, assessment to measure the emotional and behavioral needs of high-risk youth.

## Method

### Setting and Participants

The study was conducted at a residential group care facility that serves over 500 girls and boys in 70 family-style group homes in a large Midwestern city. The agency has been using an adaptation of the Teaching Family Model (TFM; Davis & Daly, 2003; Wolf et al., 1976) since the 1980's and employs married couples as the primary service delivery agents. This treatment-providing couple live in a family-style home with up to eight adolescent girls or boys. These primary treatment providers are referred to as *staff* for the rest of the manuscript.

Youth eligible to participate were identified with a disruptive behavior diagnosis (via a professional diagnosis, Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), or the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), were at least 10 years old, were experiencing their first admission to the agency, and were assigned to service providers participating in the study. Based on these criteria, over a two year period, 170 youth were eligible for participation and 145 (85%) had guardian consent and youth assent to participate. All recruitment and consent

procedures for youth and staff were approved by the University of Nebraska-Lincoln IRB and the agency IRB.

For this study, 52 staff rated 143 youth using the SFSS and the CBCL. Only youth for whom staff had responded to at least 80% of questionnaire items were included in the analyses. The sample of youth consisted of 63 girls and 80 boys. Seventeen youth indicated that they were Hispanic, 66 Caucasian, 43 African American, and 15 other. Age at enrollment ranged from 11 to 17 years, with a mean age of 15.7 years ( $sd = 1.28$ ). The majority of staff was younger than 30 years of age (54%), and held, at least, a bachelor's or associate degree (86.5%) and had been working at the agency between 1 and 6 years (71.1%).

## Procedures

Staff members were asked to complete the SFSS as a part of an assessment battery for a larger, longitudinal study (Duppong Hurley, Stevens, & Pick, 2012) on the implementation fidelity the treatment model used in this residential facility. Staff completed the SFSS, via an online or paper-and-pencil method, approximately one month after youth were admitted to care.

## Measures

**SFSS**—Since the SFSS was developed for use in outpatient settings, we worked with the developers to create a slightly modified version for use in 24/7 residential care settings. The two most salient changes were that we: (1) used a 3-point Likert-type scale (1 = *Never*, 2 = *Sometimes*, 3 = *Very Often*) instead of the 5-point scale used in outpatient settings (1 = *Never*, 2 = *Hardly Ever*, 3 = *Sometimes*, 4 = *Often*, 5 = *Very Often*), and (2) modified the wording of the items to correspond to a residential care setting (e.g., removed the reference to the ‘last session’, changed ‘clinician’ to ‘staff’, etc.). Raw SFSS scores were converted to standard scores (for the original 5-point scale) using a linking function derived from an item response theory model fit with data from a large-scale field testing study in outpatient settings (Bickman et al., 2010). The standard score for the outpatient psychometric sample has a mean of 50 and a standard deviation of 10. Standard scores greater than 43 represent *elevated problem severity* (medium and high severity), and scores less than 43 represent *low problem severity* (Bickman et al., 2010). Note that, since the SFSS was normed on only clinical samples, the SFSS scores cannot be used to identify individuals with normal versus clinical problem severity.

**CBCL**—The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) was used to evaluate the convergent validity and the usefulness of the SFSS scores in identifying emotional and behavioral problems. We selected the CBCL because it has strong psychometric properties (Achenbach & Rescorla, 2001), established clinical applications (Kazdin, 1994), and is currently in wide use across children's mental health and child welfare settings. The CBCL consists of 113 items, each rated on a 3-point Likert-type scale (0 = *not true*; 1 = *somewhat or somewhat true*; 2 = *very true or often true*), that represent two broad band scales, internalizing and externalizing problems, which are combined to form a total problems raw score. The raw score is then transformed to a standard  $t$ -score that has normed cutoffs for clinical (> 63), borderline (60 – 63), and normal problem severity (<

60). We considered clinical and borderline severity categories to represent *elevated problem severity*.

### Analysis Plan

Data analysis consisted of two primary sets of analyses: (a) correlations were computed between the SFSS and CBCL for establishing convergent validity, and (b) binary classification analyses for assessing the comparability of the SFSS and CBCL classifications. SPSS v21 was used to compute the zero-order correlations between the SFSS Total Standard Score and the CBCL Total T-Score, and between the SFSS broad-band scales (externalizing and internalizing problem severity) and the respective CBCL broad-band scales. SPSS was also used to compute the binary classification analyses. All statistical tests were evaluated for significance at the .05 alpha level.

**Binary classification analyses**—To examine the usefulness of the SFSS to identify the severity of behavioral and emotional problems, we used binary classification analysis to evaluate the relationship between two dichotomous variables, a *predicted status* (in this case, SFSS predicted normal vs. elevated severity) and a *known status* (CBCL predicted normal vs. elevated severity). Each scale was analyzed separately and classifications were evaluated for only the data collected at admission into care (within 6 weeks of admission). We analyzed these dichotomous variables with SPSS in order to calculate several indicators of classification quality including sensitivity and specificity, area under the receiver operating characteristic curve (AUROC), as well as positive and negative likelihood ratios. Multiple diagnostic indicators were reported because each one provides a different perspective or a re-expression of the classification quality of the SFSS.

The sensitivity index represents the “proportion of cases in which a disorder is detected when it is in fact present” (AERA, APA & NCME, 1999, p. 182). For the SFSS, sensitivity is the probability of an elevated problem severity test result on the SFSS given that the youth has elevated severity on the CBCL. Sensitivity can also be thought of, more simply, as the proportion of true positive classifications by the SFSS. Conversely, specificity is “the proportion of cases for which a diagnosis of disorder is rejected when rejection is warranted” (AERA, APA & NCME, 1999, p. 182). In the present study, specificity is the proportion of the sample with normal problem severity on the CBCL that is classified with normal severity on the SFSS. That is, the proportion of true negative classifications by the SFSS. If all youth were to be classified correctly by the SFSS, the sensitivity and specificity would both be 1 – 100% true positive and 100% true negative cases. Various guidelines for acceptable values have been proposed ranging from values .70 (Wood, Flowers, Meyer, & Hill, 2002) to values .90 (Johnson, Jenkins, Petscher, & Catts, 2009).

The area under the receiver operating characteristic curve (AUROC) represents the relationship between the sensitivity and specificity of an assessment. More specifically, the receiver operating characteristic curve is a plot of sensitivity by [1-specificity] across the full range of specificity. The AUROC has several interpretations centered on how well the test scores discriminate true cases from negative cases. The first interpretation is that AUROC represents, for random pair of youth (one with a CBCL score indicating elevated severity

and the other with a CBCL indicating normal severity), the likelihood that the youth with elevated severity would have a higher SFSS score than the other youth. Another interpretation is that the AUROC represents “the average value of sensitivity for all possible values of specificity” (Park, Goo, & Jo, 2004, p. 13). AUROC values greater than .90 are considered excellent, .80 to .89 are good, .70 to .79 are fair, and lesser values are poor (Compton, Fuchs, Fuchs, & Bryant, 2006). A test that lacks discriminating efficacy (i.e., classifications are no better than chance) would have an AUROC value equal to .5; therefore, this is the null hypothesis used to test the statistical significance of the observed AUROC value.

Positive likelihood ratio ( $LR^+$ ) represents a ratio of true positives to false negatives. More specifically,  $LR^+$  compares the likelihood that a youth with elevated problem severity would be classified as such by the SFSS to the likelihood that a youth with normal severity would be classified with elevated severity by the SFSS. In other words, it is the ratio of sensitivity to  $[1 - \text{specificity}]$ . Higher  $LR^+$  values indicate a more discriminating and predictive test (Cook, 2007), and specifically, one that has a low Type I error rate (i.e., false positives). Here, for example, a  $LR^+$  value of 1.3 would indicate that youth with elevated severity are 1.3 times more likely to be classified with elevated severity by the SFSS compared to youth with normal severity.

Negative likelihood ratio ( $LR^-$ ) is the ratio of false positives to true negatives. In the present analysis,  $LR^-$  compares the likelihood that youth with elevated severity would be classified with normal problem severity by the SFSS compared to the likelihood that youth with normal severity would be classified as such by the SFSS.  $LR^-$  values much lower than 1.0 are desirable because they indicate that youth with elevated severity are unlikely to be classified with normal severity by the SFSS; that is, a small value represents a test with a low Type II error rate (i.e., false negatives). For example, a  $LR^-$  of .50 would indicate that youth with normal severity are twice as likely to be classified with normal severity compared to elevated severity by the SFSS.

In addition to the indexes listed above, the Youden Index (Youden, 1950) was examined for each of the three AUROC analyses. The Youden Index ( $J$ ) is equal to  $[\text{Sensitivity} + \text{Specificity}] - 1$ . The index provides a means for comparing different threshold values so that an optimal threshold can be identified (Powers, 2011; Youden, 1950).

## Results

### Convergent Validity

The convergent validity was assessed using zero-order correlations between the SFSS scales and the CBCL scales. A large positive value indicates that the SFSS scale yields a similar pattern of scores as the CBCL scale. The correlations between total problem scales ( $r = .86$ ), externalizing scales ( $r = .78$ ) and internalizing scales ( $r = .83$ ) were all large relative to generic guidelines ( $>.5$ ; Cohen, 1988) and in line with prior research on the SFSS (Athay et al., 2012). All correlations were statistically significant at the .001 alpha level.

## Binary Classification Analyses

The results of the binary classification analyses are reported in Table 1. The sensitivity of the Total Problems scale was .70 which can be considered fair, but not quite at the acceptable level of .80 (Committee on Children with Disabilities, 2001). The sensitivity of the Total Problems scale indicates that the test has a higher than ideal Type II error rate (false negatives). Seventy percent of youth with elevated severity were correctly classified by the SFSS, so 30% of youth elevated severity were classified with normal severity by the SFSS. The specificity was .93 which is considered excellent and indicates that the SFSS does not tend to classify youth with normal severity as exhibiting elevated severity – in fact 93% of youth with normal severity were correctly classified by the SFSS. The sensitivity and specificity of the Externalizing Problems and Internalizing Problems scales were similar to the Total Problems scale (Table 1).

The overall classification accuracy (i.e., the correct classifications divided by the total number of classifications) was 84%, 77%, and 87% for the Total Problems, Externalizing Problems, and Internalizing Problems scales, respectively. The area under the receiver operating characteristic curve (AUROC) was significantly different from the null hypothesis (AUROC = .5) at the .01 alpha level for the Total Problems (.91), Externalizing Problems (.85), and Internalizing Problems (.92) scales. The AUROC values for the total scale and the subscales indicate that these scores can accurately discriminate youth with elevated problem severity at an acceptable level given that the AUROC value represents the average sensitivity across all possible values of specificity (Park, Goo, & Jo, 2004). AUROC values that are considerably higher than the observed sensitivity values, tend to indicate that the threshold for elevated severity is not optimized for this sample. In other words, there is a different threshold that would increase the sensitivity of the test while maintaining an acceptable level of specificity.

The Youden Index for the SFSS total scale was maximized ( $J = .66$ ) at the score of 40 which produced sensitivity and specificity values of .83; for the externalizing subscale ( $J = .56$ ) at the score of 38 which produced a sensitivity value of .81 and a specificity value of .76; for the internalizing subscale ( $J = .64$ ) at the score of 39 which produced a sensitivity value of .83 and a specificity value of .82.

The  $LR^+$  values were considerably large ranging from 10.0 (Total Problems scale) to 4.67 (Externalizing Problems scale) indicating that youth with elevated severity are 4.6 to 10 times more likely to be classified with elevated severity compared to youth with normal severity – this is another indicator of low Type I errors. In addition, the  $LR^-$  values were rather small (0.32, 0.35, and 0.23, respectively) indicating that youth with elevated severity were 2.8 to 4.2 times less likely to be classified with normal severity compared to youth with normal severity. While the  $LR^-$  values deviate substantially from 1.0, the  $LR^+$  values deviate much more from 1.0 indicating that the SFSS tends to have a higher rate of Type II errors (false negatives) compared to Type I errors (false positives).

## Discussion

The goal of this study was to examine the convergent validity and classification quality of the SFSS, a measure of youth behavior problem severity. Overall, the findings suggest that the SFSS may represent an acceptable alternative to longer measures of behavioral and emotional problem severity. The convergent validity of the SFSS with the CBCL was acceptable, and the ability of the SFSS to classify youth with elevated emotional and behavioral problems versus normal problem severity was marginally acceptable. A synthesis of the results indicates that the 24-item SFSS seems to be a promising measure for assessing behavioral and emotional problem severity for youth, in particular those with disruptive behavior issues receiving treatment in residential care. However, users of the SFSS should be aware of the limitations of the assessment in residential care groups with primarily high externalizing severity and relatively low internalizing problem severity. In addition, practitioners need to be cognizant regarding the purpose of the assessment and inferences that are to be made when using the SFSS. For example, if practitioners are primarily concerned about making diagnostic decisions and basing service planning around these decisions, then we recommend that the CBCL be used. On the other hand, if practitioners are primarily interested in a progress monitoring tool or an assessment of severity for internal evaluation purposes, then the SFSS seems like a very reasonable assessment to use.

The majority of the diagnostic quality indicators were acceptable; however, the sensitivity of the test was rather low (.70, .70, and .79, for the total, externalizing and internalizing scales, respectively) indicating that the test results in a high level of Type II errors (false negatives). That is, the SFSS misclassified up to 30% of the youth who had elevated levels of behavior or emotional problems. That is a notable limitation of the SFSS. For service planning the Type II error rate presents a significant drawback for practitioners as the SFSS might not indicate the need for additional services for youth with elevated problem severity. However, as the AUROC values indicate, the SFSS scores are, in fact, discriminating, but the currently used threshold does not seem to be set at the optimal score (i.e., the best balance between sensitivity and specificity). We see three possible reasons for this discrepancy between sensitivity and the AUROC: (a) the threshold for elevated severity was developed based solely on clinical samples and therefore cannot accurately distinguish between the presence or absence of 'clinical' level severity (which differs from the CBCL since it was normed with non-clinical samples), (b) the linking function used to derive standard scores from the 3-point scale needs revision, or (c) the CBCL is a substantially more sensitive instrument for discriminating between youth with borderline severity and youth with normal severity. To address the optimal cut score, the Youden Index was calculated for a series of possible thresholds until the optimal score was found; however, the threshold values reported in this paper should be viewed with caution since they are based on a single, relatively small sample. Although the sensitivity of the SFSS scores is relatively low, the specificity is high, and the  $LR^+$  is large indicating that the test results in a low Type I error rate (false positives) compared to true negative classifications.

## Limitations

This study is limited in focusing on a single provider of residential care for youth; future studies need to replicated these findings with other providers. Further, the sample size for this study was relatively small; however, given the small samples size of most group-homes, this represents a large sample within the field of residential care, and currently provides the most information on the SFSS in residential care to date. Another limitation of the study was that only youth with disruptive behavior diagnosis were included. While this helped to lessen variance due to diagnosis issues, future studies need to see how the SFSS performs with youth in residential settings with other diagnoses.

## Implications and Future Research

These findings suggest that the SFSS can be used with high-needs youth with reasonable confidence that the measurement and classification properties are similar to those of most established, larger assessments. While the SFSS scores were similar to CBCL scores for classifying elevated and normal problem severity, more research is needed to investigate the classification capabilities of the SFSS with other respondents (e.g. parents, youth self-report) as well as in other treatment settings. Future studies are also needed to examine how the SFSS performs across different subgroups of youth, such as different diagnostic categories. More longitudinal studies are also needed to see if SFSS scores are predictive of functional outcomes, such as youth discharging from care or returning for another placement.

Given that the measure is considerably shorter than the CBCL, it is far more feasible to administer frequently during care to monitor treatment progress for youth. This is essential, as there is a paucity of measures that have strong psychometric properties, are brief enough for routine-progress monitoring, and can be completed by multiple respondents that have also been successfully tested with youth in out-of-home treatment. To this end, research on sensitivity to change is needed to help establish the usefulness of the SFSS for progress monitoring. Moreover, the SFSS is unique as it has been incorporated into the Contextualized Feedback System (Bickman, 2008), which is web-based client management software that tracks client progress and treatment process measures while providing real-time tracking reports. Being able to administer measures quickly, efficiently, and use them for real-time data decision making is a key component to improving the quality of service delivery (Bickman, 2012; Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Chorpita, Reise, Weisz, Grubbs, Becker, & Krull, 2010; Ebesutani, C., Berstein, A., Chorpita, B., & Weisz, 2012).

In sum, this study extends the research on the SFSS for use with youth with a disruptive behavior diagnosis, including its ability to classify youth with elevated or normal emotional or behavioral problem severity. It provides support for the use of the SFSS with the high-need youth, affording a much needed brief assessment tool that can be used for routine progress monitoring.

## Acknowledgments

The research reported herein was supported, in part, by the National Institute of Mental Health through Grant R34MH080941 and by the Institute of Education Sciences, U.S. Department of Education, through Grant

R324B110001 to the University of Nebraska-Lincoln. The opinions expressed are those of the authors and do not represent views of the National Institute of Mental Health or the U.S. Department of Education.

## References

- Achenbach, TM.; Rescorla, L. ASEBA School-Age Forms & Profiles. Research Center for Children, Youth, and Families; Burlington, VT: 2001.
- Athay MM, Riemer M, Bickman L. The Symptoms and Functioning Severity Scale (SFSS): Psychometric evaluation and discrepancies among youth, caregiver, and clinician ratings over time. *Administration and Policy in Mental Health and Mental Health Services Research*. 2012; 39:13–29. [PubMed: 22407556]
- Arseneault L, Kim-Cohen J, Taylor A, Caspi A, Moffitt TE. Psychometric evaluation of 5- and 7-year-old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology*. 2005; 33:537–550. [PubMed: 16195949]
- Bickman L. A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Association of Child and Adolescent Psychiatry*. 2008; 47:1114–1119.
- Bickman L. Why can't health services be more like baseball?. *Administration and Policy in Mental Health and Mental Health Services Research*. 2012; 39(1):1–2. [PubMed: 22407560]
- Bickman, L.; Athay, MM.; Riemer, M.; Lambert, EW.; Kelley, SD.; Breda, C.; Tempesti, T.; Dew-Reeves, SE.; Brannan, AM.; de Andrade, AR., editors. *Manual of the Peabody Treatment Progress Battery*, 2nd ed. [Electronic version]. Vanderbilt University; Nashville, TN: 2010. <http://peabody.vanderbilt.edu/ptpb/>
- Bickman L, Kelley SD, Breda C, de Andrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services*. 2011; 62(11):1423–1429. [PubMed: 22193788]
- Bradley R, Doolittle J, Bartolotta R. Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education*. 2008; 17:4–23.
- Brown JD, Hamilton M, Natzke B, Ireys HT, Gillingham M. Use of out-of-home care among a statewide population of children and youth enrolled in medicaid. *Journal of Child and Family Studies*. 2011; 20:48–56.
- Charles JM, Bywater T, Edwards RT. Parenting interventions: a systematic review of the economic evidence. *Child: Care, Health and Development*. 2011; 37:462–474.
- Chorpita BF, Reise S, Weisz JR, Grubbs K, Becker KD, Krull JL. Evaluation of the Brief Problem Checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*. 2010; 78(4):526–536. [PubMed: 20658809]
- Ebesutani C, Berstein A, Chorpita B, Weisz JR. A transportable assessment protocol for prescribing youth psychological treatments in real-world settings: Reducing assessment burden via self-report scales. *Psychological Assessment*. 2012; 24(1):141–155. [PubMed: 21859220]
- Epstein, MH. *Behavioral and Emotional Rating Scale, Second Edition Examiner's Manual*. Pro-Ed; Austin, TX: 2002.
- Goodman R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*. 1997; 38:581–586. [PubMed: 9255702]
- Kolko DJ, Pardini DA. ODD dimensions, ADHD, and callous–unemotional traits as predictors of treatment response in children with disruptive behavior disorders. *Journal of Abnormal Psychology*. 2010; 119:713. [PubMed: 21090875]
- McKinney C, Morse M. Assessment of disruptive behavior disorders: Tools and recommendations. *Professional Psychology: Research and Practice*. 2012; 43:641.
- Reynolds, CR.; Kamphaus, RW. *Behavior Assessment System for Children – Second Edition Manual*. Pearson: Minneapolis, MN: 2004.
- Short RJ, Shapiro SK. Conduct disorders: A framework for understanding and intervention in schools and communities. *School Psychology Review*. 1993; 22:362–375.

- Trentacosta CJ, Hyde LW, Goodlett BD, Shaw DS. Longitudinal prediction of disruptive behavior disorders in adolescent males from multiple risk domains. *Child Psychiatry & Human Development*. 2013; 44:561–572. [PubMed: 23239427]
- Webster-Stratton C. Strategies for helping early school-aged children with oppositional defiant and conduct disorders: The importance of home-school partnerships. *School Psychology Review*. 1993; 22:437–457.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

## Binary Classification Results

|                        | <b>Sensitivity</b> | <b>Specificity</b> | <b>Acc</b> | <b>AUROC</b> | <b>LR<sup>+</sup></b> | <b>LR<sup>-</sup></b> |
|------------------------|--------------------|--------------------|------------|--------------|-----------------------|-----------------------|
| Total Problems         | 0.70               | 0.93               | 0.84       | 0.91         | 10.00                 | 0.32                  |
| Externalizing Problems | 0.70               | 0.85               | 0.77       | 0.85         | 4.67                  | 0.35                  |
| Internalizing Problems | 0.79               | 0.89               | 0.87       | 0.92         | 7.53                  | 0.23                  |

Acc = overall classification accuracy

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript