

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Special Education and Communication Disorders
Faculty Publications

Department of Special Education and
Communication Disorders

2015

Psychometric Evaluation of the Symptoms and Functioning Severity Scale (SFSS) Short Forms with Out-of-Home Care Youth

Thomas J. Gross

University of Nebraska-Lincoln, tjgross44@gmail.com

Kristin Duppong-Hurley

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

Matthew C. Lambert

University of Nebraska-Lincoln, mlambert2@unl.edu

Michael Epstein

University of Nebraska - Lincoln, mepstein1@unl.edu

Amy L. Stevens

National Research Institute at Father Flanagan's Boys' Home, amy.stevens@boystown.org

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



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

Gross, Thomas J.; Duppong-Hurley, Kristin; Lambert, Matthew C.; Epstein, Michael; and Stevens, Amy L., "Psychometric Evaluation of the Symptoms and Functioning Severity Scale (SFSS) Short Forms with Out-of-Home Care Youth" (2015). *Special Education and Communication Disorders Faculty Publications*. 117.

<http://digitalcommons.unl.edu/specedfacpub/117>

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:

Child Youth Care Forum. 2015 April 1; 44(2): 239–249. doi:10.1007/s10566-014-9280-z.

© Springer Science+Business Media New York 2014. Used by permission.

Psychometric Evaluation of the Symptoms and Functioning Severity Scale (SFSS) Short Forms with Out-of-Home Care Youth

Thomas J. Gross,

University of Nebraska – Lincoln, 213 Barkley Memorial Center, Lincoln, NE 68583,
402-472-5484

Kristin Duppong Hurley,

University of Nebraska – Lincoln, 247E Barkley Memorial Center, Lincoln, NE 68583,
402-472-5501

Matthew C. Lambert,

University of Nebraska – Lincoln, 273 Barkley Memorial Center, Lincoln, NE 68583,
402-472-5487

Michael H. Epstein, and

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

Amy L. Stevens

National Research Institute at Father Flanagan's Boys' Home, 100 Crawford Drive, Boys Town,
NE 68010

Thomas J. Gross: tjgross44@gmail.com; Kristin Duppong Hurley: kristin.hurley@unl.edu; Matthew C. Lambert: mlambert2@unl.edu; Michael H. Epstein: mepstein1@unl.edu; Amy L. Stevens: amy.stevens@boystown.org

Abstract

BACKGROUND—There is a need for brief progress monitoring measures of behavioral and emotional symptoms for youth in out-of-home care. The Symptoms and Functioning Severity Scale (SFSS; Bickman et al., 2010) is one measure that has clinician and youth short forms (SFSS-SFs); however, the psychometric soundness of the SFSS-SFs with youth in out-of-home care has yet to be examined.

OBJECTIVE—The objective was to determine if the psychometric characteristics of the clinician and youth SFSS-SFs are viable for use in out-of-home care programs.

METHODS—The participants included 143 youth receiving residential treatment and 52 direct care residential staff. The current study assessed internal consistency and alternate forms reliability for SFSS-SFs for youth in a residential care setting. Further, a binary classification test was completed to determine if the SFSS-SFs similarly classified youth as the SFSS full version for low- and elevated-severity.

RESULTS—The internal consistency for the clinician and youth SFSS-SFs was adequate ($\alpha = .75$ to $.82$) as was the parallel forms reliability ($r = .85$ to $.97$). The sensitivity (0.80 to 0.95), specificity (0.88 to 0.97), and overall accuracy (0.89 to 0.93) for differentiating low and elevated symptom severity was acceptable.

CONCLUSIONS—The clinician and youth SFSS-SFs have acceptable psychometrics and may be beneficial for progress monitoring and additional research should clarify their potential for progress monitoring of youth in out-of-home programs.

Youth in out-of-home care tend to exhibit persistent internalizing and externalizing behavioral and emotional problems (Baker, Archer, & Curtis, 2007; Butler & Richard, 2013; Cuthbert et al., 2011). They frequently meet criteria for diagnoses of conduct disorder, mood and anxiety disorders, and adjustment disorders (Connor, Doerfler, Toscano, Volungis, & Steingard, 2004; Turner & Macdonald, 2011). Out-of-home care is typically characterized by facilities that provide well controlled and intensive treatment, which includes supervision, structured daily living, and psychoeducational interventions (Brown Hamilton, Natzke, Ireys, & Gillingham, 2011). These programs demonstrate effectiveness when they show decreased behavioral and emotional problems (Robst, Armstrong, & Dollard, 2011); it is typically preferred to use assessment methods that are quick, time efficient and cost effective (Larzelere, Daly, Davis, Chmelka, & Handwerk, 2004). There are many symptom assessment tools available, but this does not mean they are suited to assess symptoms for all problems or across all treatment settings (Kazdin, 2005). Unfortunately, the current state of psychometrically sound, concise, and repeatable psychological symptom assessment for youths in out-of-home care is insufficient (Chambers, Saunders, New, Williams, & Stachurska, 2010). For example, the Assessment Checklist for Children (ACC; Tarren-Sweeney, 2007) and Assessment Checklist for Adolescents (ACA; Tarren-Sweeney, 2013a) are available in long forms with over 100 items each or short forms with 20 items each (Tarren-Sweeney, 2013b). These short forms may be too lengthy for frequent assessments of youth functioning and different questionnaires are required for different age ranges (4 to 11 years and 12 to 17 years; Tarren-Sweeney, 2013b), which may become cumbersome.

Progress monitoring measures are continuous, routine assessments of client change that can help identify a client's response to treatment. The purpose of progress monitoring is to inform clinical judgments regarding client progress and functioning (Overington & Ionita, 2012) and aide in decision making regarding changes to a specific intervention (Andrews & Page, 2005; Gresham et al., 2010). In general, six common criteria for progress monitoring measures are: (a) reliability and social validity, (b) sensitivity to change, (c) a format that allows routine administration, (d) assessment of overall difficulties and functioning, (e) easy administration, and (f) application across needs and settings (Gresham et al., 2010; Kazdin, 2005; Overington & Ionita, 2012). Many in the field of youth behavioral and emotional treatment are unsure of the best method for progress monitoring (Gresham et al., 2010; Kazdin, 2005). Rating scales are one method that is widely used to assess behavior because they reliably sample behaviors through a cost-effective means, but their single biggest limiting factor is length (Volpe & Gadow, 2010). It has been assumed that a large number of items are needed to adequately represent overall difficulties and functioning. However, some have demonstrated that rating scales may be shortened and continue to have adequate

internal consistency, while sufficiently representing overall difficulties and functioning (e.g., Gresham et al., 2010; Volpe & Gadow, 2010).

The Symptoms and Functioning Severity Scale full version (SFSS; Bickman et al., 2010) is an instrument designed to assess emotional and behavioral problems in youth aged 11 to 18 years. The SFSS is a general measure of youth problems and useful in noting changes in symptoms and functioning for the purpose of progress monitoring. The SFSS was designed with the purpose of representing the range of severity level for typically problems youths may experience. That is, reduced symptom severity implies increased personal functioning and increased symptom severity represents decreased functioning (Bickman et al., 2010). The Diagnostic and Statistical Manual of Mental Disorders –Text Revision (DSM-IV-TR) was used as a guide for item development to ensure key characteristics of internalizing and externalizing disorders were represented. However, its purpose is not as a diagnostic instrument, but as a global indicator that provides a continuous metric for symptom severity and functionality (Bickman et al., 2010).

The SFSS has clinical utility because it is relatively brief, sensitive to change, and provides clinically relevant information across different settings and raters (Athay, Riemer, & Bickman, 2012). The SFSS clinician version contains 27 items, of which 24 are scored to yield a Total Problems score. The 24 items are distributed among two subscales, Externalizing Problems (14 items) and Internalizing Problems (10 items). Internal consistency was found to be acceptable for the Total Problems ($\alpha = .94$), Externalizing Problems ($\alpha = .93$), and Internalizing Problems scales ($\alpha = .89$; Bickman et al., 2010). Moreover, the SFSS clinician version substantially correlates to other measures of youth behavioral and emotional symptoms, such as the Child Behavior Checklist (CBCL; $r = .86$) and Strengths and Difficulties Questionnaire (SDQ; $r = .71$ to $.79$; Athay et al., 2012), which are well-designed instruments for detecting behavioral problems or psychiatric diagnoses (Goodman, 1997; Kazdin, 1994). The SFSS youth version contains 26 items, of which 24 are scored to yield a Total Problems score. The 24 items are distributed into Externalizing Problems and Internalizing Problems in the same manner as the clinician version. Internal consistency was reported to be acceptable for the Total Problems ($\alpha = .92$), Externalizing Problems ($\alpha = .89$), and Internalizing Problems scales ($\alpha = .88$; Bickman et al., 2010).

Considering the time constraints of administering even a 26-item measure repeatedly in out-of-home care settings, the SFSS full version is likely too long for routine use despite its promising psychometric properties. However, 14-item short-forms of the SFSS include two abridged, non-redundant forms that assess overall symptom severity designed for repeated use (Bickman et al., 2010). The SFSS Short Form A (SFSS-SF-A) and SFSS Short Form B (SFSS-SF-B) include clinician and youth versions, like the SFSS full version, and the SFSS authors recommend alternating administrations of the short forms every other week for progress monitoring. The SFSS-SF-A and -B are composed of separate, but parallel items from the SFSS; and they yield a Total score, but do not provide Internalizing and Externalizing problem scale scores. The SFSS-A and -B were created by matching items on content area and psychometric similarity, and then one half of the items were included in SFSS-A and the other half were included on SFSS-B (Bickman et al., 2010).

The clinician version of the SFSS-SF-A and SFSS-SF-B each contain 14 items from the SFSS full version, of which 12 items provide a Total Problems score, and both forms have acceptable internal consistency ($\alpha = .86$ to $.87$; Bickman et al., 2010). The youth version of the SFSS-SF-A and SFSS-SF-B each contain 13 different items from the SFSS, of which 12 items are scored to provide a Total Problems score, and both forms have acceptable internal consistency ($\alpha = .84$ to $.86$; Bickman et al., 2010). Additionally, the clinician and youth SFSS-SFs have similar means, standard deviations, and score distributions to their full version counterparts (Athay et al., 2012; Bickman et al., 2010).

It is encouraging that the SFSS has short forms which may provide quick assessments of symptoms for progress monitoring purposes; however, the validity of the SFSS-SFs was only tested with outpatient participants (e.g., Athay et al., 2012; Bickman et al., 2010). It is widely accepted that in order for an instrument to be used with a population, which is substantially different from the normative sample, the psychometric qualities should be reassessed (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing [AERA, APA & NCME], 1999). Moreover, the SFSS authors may have made further modifications to the SFSS items in an attempt to make it more compatible with measures already used in out-of-home care (e.g., reducing the 5-point scale to a 3-point scale, altering items' words). An initial study of the modified clinician and youth SFSS full versions with residential staff and youth found acceptable internal consistency, and acceptable classification accuracy between low- and elevated-severity groups when compared to the CBCL (Duppong Hurley, Lambert, & Stevens, 2014; Lambert, Duppong Hurley, Gross, Epstein, & Stevens, 2014). No studies have examined the internal consistency or parallel forms reliability of the SFSS-SF-A and -B in their current modified formats. Further, the short forms have not been assessed to determine if they similarly classify youth as the SFSS full version for severity level.

The purpose of this study was to assess the internal consistency and parallel forms reliability in a residential setting. The classification accuracy (i.e., classification similarity to the full version) of the clinician and youth SFSS-SF-A and -B was also assessed. The objective was to examine the psychometric characteristics of the clinician and youth SFSS-SFs. That is, determine if they are reliable measures, which reflect overall symptom severity and functioning similar to the long form, and have application in out-of-home settings. Thus, they may demonstrate potential as a viable measure of youth behavior for out-of-home care programs and meet three of the six common criteria for progress monitoring measures.

Method

Participants and Setting

The participants included 143 youths receiving residential treatment and 52 direct care residential staff. The youths consisted of 63 girls and 80 boys, and were between 11 to 17 years ($M = 15.7$, $SD = 1.28$). Youth participants indicated 12% Hispanic, 46% Caucasian, 30% African-American, and 10% other racial/ethnic heritages. The staff worked at the residential facility between 1 to 6 years, and a large proportion held an associate or

bachelor's degree (87%). The youths were identified with a disruptive behavior disorder through professional diagnosis, Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), or the CBCL (Achenbach & Rescorla, 2001); they were 10 years of age or older, received their first admission to the residential program, and were in the care of residential staff participating in the study.

The study took place at a residential group care facility in a large Midwestern city, which adapted the Teaching Family Model (TFM; Davis & Daly, 2003; Wolf et al., 1976) over 30 years ago. The TFM stresses naturalistic family-like practices within structured daily routines for youth to learn social and academic skills as they are gradually introduced to typical environments and activities to practice these skills (Fixsen & Blase, 1993; Kirigin, 2001; Wolf, Kirigin, Fixsen, Blasé, & Braukmann, 1995). The facility serves over 500 youths in 70 family-style group homes. The residential facility employs married couples (i.e., family teachers) as the primary service-delivery agents, who live in family-style homes with up to eight adolescent girls or boys. All recruitment and consent procedures for youth and staff were approved by the University of Nebraska-Lincoln IRB and the agency IRB.

Measures

SFSS—The SFSS was developed for use with outpatient populations and uses a five-point Likert-like scale (1 = *Never*, 2 = *Hardly Ever*, 3 = *Sometimes*, 4 = *Often*, 5 = *Very Often*). The researchers worked with the SFSS developers to modify it for use in residential care settings. The wording of items was changed to reflect residential care (e.g., referenced the “last two weeks” instead of since the previous “session”), and the five-point Likert-like scale was changed to a three-point Likert-like scale (1 = *Never*, 2 = *Sometimes*, 3 = *Very Often*). It should be noted that the decision to use a 3-point rating scale instead of a 5-point for out-of-home settings was made by the developers of SFSS and their desire to use a 3-point rating scale was respected by the research team.

The clinician SFSS-SF-A consisted of 14 items and the youth SFSS-SF-A consisted of 13 items, of which the same 12 items are used to obtain Total Problems score. The clinician and youth SFSS-SF-B forms are structured in the same manner and use separate, parallel items. For example, the SFSS-SF-A items “Hard time waiting turn” and “Feel nervous/shy” correspond respectively to SFSS-SF-B items “Interrupted others” and “Worry about a lot of things”(see Bickman et al., 2010). The short forms contain internalizing and externalizing items; however, they yield only a Total severity score.

Raw SFSS-SF scores were converted to standard scores (reflecting 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 and the converted SFSS-SF scores assures similar score distribution. Total scores for the clinician short forms can range from 24 to 79. Total scores below 43 indicate low severity, scores from 43 to 57 indicate medium severity, and scores above 57 indicate high severity (Bickman et al., 2010). Total scores for the youth short forms can range from 31 to 81. Total scores below 42 indicate low severity, scores from 42 to 56 indicate medium severity, and scores above 56 indicate high severity (Bickman et al., 2010). The SFSS-SF

was developed to use the same severity level scores for the 3-point rating scale as the 5-point rating scale.

CBCL—The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) was used to evaluate the convergent validity (i.e., the correspondence of total problems) of the SFSS scores. The CBCL has strong psychometric properties (Achenbach & Rescorla, 2001), established clinical applications (Kazdin, 1994), and is widely used across multiple child treatment 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).

Procedures

Family teachers and youth completed the SFSS as a part of an assessment battery for a larger, longitudinal study regarding common therapeutic process factors and the treatment fidelity of the model used in the residential facility (for details regarding recruitment and participation rates see Duppong Hurley, Lambert, Van Ryzin, Sullivan, & Stevens, 2013). Family-teachers completed the SFSS either on-line or by paper-and-pencil one month after a youth was admitted to residential care. The youth completed the SFSS by paper-and-pencil within one week of admission and after receiving guardian consent. The SFSS-SF scores were calculated from the responses on the SFSS full versions.

Analyses

Cronbach's alpha coefficients were computed for each of the SFSS-SFs to determine internal consistency. Pearson's *r* coefficients were calculated between respective SFSS-SFs to assess parallel forms reliability, and across full versions and their respective short forms to assess concurrent validity. Concurrent validity between short and full forms are necessary to assure the items in each short form reflect the same overall construct as represented in the previous (Cicchetti, 1994) or long form (Straus & Douglas, 2004). Also, clinician SFSS-SF scores were correlated with the CBCL Total Problems T-scores from residential staff because both assessments are continuous measures of overall youth symptoms and functioning. The Total Scores from all the SFSS forms were used to divide youths into two groups, low- and elevated-severity (medium severity scores or higher). A binary classification test was completed to determine the sensitivity, specificity, and overall accuracy of the short forms compared to the long form. This allows for the determination that all short forms classify youth similarly to the full forms. Sensitivity is the proportion of true positive classifications, or youth correctly classified as low- or elevated-severity. Specificity is the proportion of true negative classifications, or the tendency to correctly not classify youth as low-severity when they are elevated-severity or vice versa (AERA, APA & NCME, 1999). Classification accuracy is the correct classifications divided by the total number of classifications, or correct youth classifications divided by all correct and incorrect classifications. Comprehensive analyses of the clinician and youth SFSS full versions to CBCL scores and classifications are available in Lambert et al. (2014).

Results

Internal Consistency and Alternate Forms Reliability

Alpha coefficients for the clinician SFSS-SF-A and -B were .78 and .82, respectively; and .77 and .75 for the youth SFSS-A and -B, respectively. Alpha coefficients for the two staff short forms did not differ statistically ($w = 0.81, p = .12$) nor did the coefficients for the youth short forms ($w = 0.92, p = .31$; Feldt, 1969). Nunnally (1978) proposed a generally acceptable alpha level of .80, and Cicchetti (1994) provided guidelines for clinical utility of below .40 as poor, between .40 to .59 as fair, .60 to .74 as good, and .75 to 1.00 as excellent. Given the current coefficient values the internal reliability of the clinician and youth short forms can be considered generally adequate and excellent for clinical use. The clinician short forms SFSS-A and -B are highly correlated with one another ($r = .85, p < .001$) as well as with the full version ($r = .96$ and $.96$, respectively). The clinician SFSS-SF-A and -B scores were highly correlated with the CBCL Total Problems scores ($r = .83$ and $.81, p < .001$, respectively). The youth short form SFSS-SF-A and -B are also highly correlated ($r = .86, p < .001$) as are each of the short forms with the full version ($r = .96$ and $.97$, respectively).

Classification Accuracy between Short and Long Forms

Table 1 contains the sensitivity, specificity, and overall accuracy of the SFSS short forms compared to the full versions. The sensitivity, specificity, and overall accuracy represent the degree of match of the short forms to the full versions; therefore, the coefficients may be interpreted as the proportion of cases correctly matched to the original full version classification for each respective form. Guidelines for acceptable values have been proposed ranging from values .70 (Wood, Flowers, Meyer, & Hill, 2002) to .90 (Johnson, Jenkins, Petscher, & Catts, 2009). Clinician and youth SFSS-SFs had acceptable sensitivity, specificity, and accuracy. The sensitivity and specificity of each short form indicated that the classification scores from the short forms align closely with the classification scores from the full versions.

Discussion

The purpose of this study was to assess the internal consistency, alternate forms reliability, and classification accuracy of the clinician and youth SFSS-SFs for use in out-of-home care. It was found that the clinician and youth SFSS-SF-A and -B had acceptable levels of internal consistency for clinical use. Scores from the clinician short forms were highly correlated with each other as were the scores from the youth short forms. The clinician and youth short forms demonstrated acceptable classification sensitivity, specificity, and accuracy between low- and elevated-severity groups.

Psychometric Assessment

Previous outpatient estimates of internal consistency for the clinician and youth short forms were greater than those found in this study. The Cronbach's α coefficients found by Bickman et al. (2010) were above the suggested cut-off of .80 (Nunnally, 1978), whereas the Cronbach's α coefficients for this study approximated the suggested level. However, when

using Cicchetti's (1994) criteria all of the short form coefficients were in the clinically excellent range of .75 to 1.00. A decrease in internal consistency coefficients was expected because separating items into parallel scales has a cumulative effect on measurement error (Henson, 2001). That is, the error associated with the of the original scale compounds with the creation of each parallel form leading to a reduction in internal consistency. Still, all short forms demonstrated acceptable internal consistency for use in out-of-home care settings. Administering the short forms as recommended by the SFSS developers could provide reliable estimates of symptom severity in residential treatment.

Previous psychometric investigations of the SFSS have not presented the intercorrelations between the respective SFSS-SFs (e.g., Athay et al., 2012; Bickman et al., 2010); however, this study examined these relationships. The correlations within the clinician and youth SFSS short forms were found to be acceptable, and the short forms had strong correlations with their respective full versions. This indicated that the parallel short forms estimated overall symptom severity in a similar manner as each other and as their respective full versions. This is important because all respective versions of the short forms are likely to yield similar scores for overall symptom severity and functioning from the same rater. The SFSS-SFs may be used interchangeably, or as a reasonable substitute for the full versions, to provide information regarding overall youth symptom severity and functioning at a momentary assessment.

The clinician and youth SFSS-SF-A and -B had acceptable agreement with the SFSS full version categorizations of low- and elevated-severity. The sensitivity and specificity of all short forms were above .70 and all but one of the short forms demonstrated sensitivity above or approximate to the more stringent threshold of .90. The sensitivity and specificity of the SFSS-SFs allow practitioners to have confidence that the SFSS-SFs will identify different levels of youth symptom severity in a similar manner as the full versions. The high specificity demonstrates that practitioners could have confidence that youth in the elevated-severity range may be at-risk for behavioral problems, if not already exhibiting behavioral symptoms. Overall, the SFSS-SFs have similar classification accuracy as the SFSS full version for differentiating low and elevated symptom severity in youth. Further, high correlations between clinician SFSS-SF-A and -B scores and the CBCL Total Problems T-scores indicated that the clinician SFSS-SFs may provide similar overall continuous ratings of problems as other, longer assessments.

SFSS Short Forms Application

There is a need for sound, concise, and repeatable psychological symptom assessment for youths in out-of-home care (Chambers et al., 2010), and the SFSS-SFs may address this need. The SFSS-SFs are brief and repeatable alternate forms that indicate overall youth symptom severity. Behavioral assessment in out-of-home care has issues with consistency across different raters from intake and discharge as well as longer assessments may become exhausting (Larzelere et al., 2001). The SFSS-SFs may ease these concerns by having the same staff member rate the same youth multiple times using a brief measure. The SFSS-SFs provide consistent and relevant information, which is a requirement of progress monitoring (McKinny & Morse, 2012).

Progress monitoring relies on brief measures to provide relevant information regarding youth behavioral and emotional functioning (Overington & Ionita, 2012) and the SFSS-SFs appear to have potential as progress monitoring measures in this regard. The SFSS-SFs could differentiate between low- and elevated-severity groups accurately which may inform clinical judgments regarding youth behavioral and emotional symptom severity. For instance, out-of-home care professionals could quickly determine if youths are experiencing low or elevated symptom severity. Further, this study found that the SFSS-SFs meet at least three of six common criteria for progress monitoring measures. The SFSS-SFs are reliable (internally consistent and reliable alternate forms), reflect overall difficulties, and may be useful across settings (i.e., out-of-home care settings and outpatient settings; Athay et al., 2012; Bickman et al., 2010).

The SFSS-SFs could be used as part of a progress monitoring system along with other assessment methods. A comprehensive picture of youth symptoms and functioning could include the SFSS-SFs along with observations or permanent products like daily report cards or point cards in instances of token economies. This complimentary information may improve clinical judgment by incorporating objective and subjective information. Moreover, the SFSS-SFs may allow for routine examinations of symptoms and functioning and the results could be synthesized with observational data to understand general daily functioning. Lengthier assessments conducted every few months could integrate this information to provide diagnostic clarity as well as adjustment of treatment goals.

Limitations and Future Directions

This study is limited in that the SFSS was used with a single provider of residential treatment. Future studies could build on this initial research to determine the SFSS-SFs' internal consistency, alternate forms reliability, and classification accuracy within other out-of-home care settings. Further, the sample size in this study was small, although it is a relatively large sample for residential care. A larger participant pool would improve confidence in the SFSS-SFs' psychometrics for youth in out-of-home care. A notable limitation of the analysis was the lack of accounting for non-independence in the youth data that arose from the nesting of youth within staff members (that is, the same staff member rated more than one youth). Although the point estimates for the correlations are unbiased, the standard errors are not, leading to a potentially inflated Type I error rate. However, as noted by some methodologists (Muthen & Satorra, 1995), ignoring nested data structures seems to be particularly problematic only when design effects (Kish, 1965) exceed a value of 2 and since many of the clusters are small (e.g., 1 to 3 youth per staff member) the resultant design effects are small. For example, the design effect for the total SFSS score was 1.08 suggesting that omitting the nested structure, in all likelihood, did not cause a significant inflation of the Type I error rate. Another limitation is that the youth in this study were only identified with disruptive behavior problems. Additional studies need to extend this line of research to out-of-home care youth with a broader range of behavioral and emotional problems. For instance, research should examine if youth with high internalizing and externalizing symptoms are rated differently on the SFSS-SFs.

Also, future studies need to assess the SFSS-SFs as an outcome and progress monitoring measure. The current study only demonstrates the adequacy at one time point, whereas future studies should determine whether the SFSS-SF-A and -B are sensitive to change over time when compared to each other and the SFSS full form. Similar sensitivity to change comparisons may be made using other behavior scales typically used out-of-home care, such as the CBCL. Moreover, the SFSS was administered in its full form and then divided into the short forms. It would be beneficial to administer the SFSS in full and short forms separately to determine if the short forms retain their psychometric properties when administered independently of the full form.

Conclusion

In conclusion, the clinician and youth SFSS-SFs are a set of measures that address the need for cost-effective and concise progress monitoring of youth in out-of-home care. The short forms had adequate internal consistency, alternate form reliability, and classification accuracy when compared to the clinician and youth SFSS full versions. The SFSS-SFs may be a beneficial component to routine progress monitoring systems and have the potential to be integrated into larger, less frequent assessment batteries. Additional research should clarify their potential for progress monitoring across youth problems and out-of-home care settings.

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, T.M.; Rescorla, L.A. *Manual for ASEBA School-Age Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families; 2001.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
- Andrews G, Page AC. Outcome measurement, outcome management and monitoring. *Australian and New Zealand Journal of Psychiatry*. 2005; 39:649–651. [PubMed: 16050918]
- 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]
- Baker AJ, Archer M, Curtis P. Youth characteristics associated with behavioral and mental health problems during the transition to residential treatment centers: the Odyssey Project population. *Child welfare*. 2007; 86:5–29.
- Bickman, L.; Athay, MM.; Riemer, M.; Lambert, EW.; Kelley, SD.; Breda, C.; Andrade, AR., editors. *Manual of the Peabody Treatment Progress Battery*. 2nd. Nashville, TN: Vanderbilt University; 2010. [Electronic version] Available from <http://peabody.vanderbilt.edu/ptpb>
- 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.

- Butler LS, Richard M. Evaluating components of residential treatment success for the most complex youth. *Residential Treatment for Children & Youth*. 2013; 30:119–130.
- Chambers MF, Saunders AM, New BD, Williams CL, Stachurska A. Assessment of children coming into care: Processes, pitfalls and partnerships. *Clinical Child Psychology and Psychiatry*. 2010; 15:511–527. [PubMed: 20923900]
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–290.
- Connor DF, Doerfler LA, Toscano PF Jr, Volungis AM, Steingard RJ. Characteristics of children and adolescents admitted to a residential treatment center. *Journal of Child and Family Studies*. 2004; 13:497–510.
- Cuthbert R, Pierre JS, Stewart SL, Cook S, Johnson AM, Leschied AW. Symptom persistence in seriously emotionally disordered children: Findings of a two-year follow-up after residential treatment. *Child & Youth Care Forum*. 2011; 40:267–280.
- Davis, J.; Daly, DL. Long-Term Residential Program training manual. 4th. Boys Town, NE: Father Flanagan's Boy's Home; 2003.
- Duppong Hurley K, Lambert MC, Stevens AS. Psychometrics of the Symptoms and Functioning Severity Scale for high-risk youth. *Journal of Emotional and Behavioral Disorders*. 2014 Advance online publication. 10.1177/1063426614535809
- Duppong Hurley K, Lambert MC, Van Ryzin M, Sullivan J, Stevens A. Therapeutic alliance between youth and staff in residential group care: Psychometrics of the therapeutic alliance quality scale. *Children and Youth Services Review*. 2013; 35:56–64. [PubMed: 23264715]
- Feldt LS. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*. 1969; 34:363–373.
- Fixsen DL, Blasé KA. Creating new realities: Program development and dissemination. *Journal of Applied Behavior Analysis*. 1993; 26:597–615. [PubMed: 8307838]
- Goodman R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*. 1997; 38:581–586. [PubMed: 9255702]
- Gresham FM, Cook CR, Collins T, Dart E, Rasetshwane K, Truelson E, Grant S. Developing a change-sensitive brief behavior rating scale as a progress monitoring tool for social behavior: An example using the Social Skills Rating System-Teacher form. *School Psychology Review*. 2010; 39:364–379.
- Henson RK. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*. 2001; 34:177–189.
- Johnson ES, Jenkins JR, Petscher Y, Catts HW. How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*. 2009; 24:174–185.10.1111/j.1540-5826.2009.00291.x
- Kazdin, AE. Methodology, design and evaluation in psychotherapy research. In: Bergin, AE.; Garfield, SL., editors. *Handbook of psychotherapy and behaviour change*. 4th. New York, NY: Wiley; 1994. p. 19-71.
- Kazdin AE. Evidence-based assessment for children and adolescents: Issues in measurement development and clinical application. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:548–558. [PubMed: 16026218]
- Kirigin KA. The Teaching-Family Model: A replicable system of care. *Residential Treatment for Children and Youth*. 2001; 18:9–110.
- Kish, L. *Survey Sampling*. New York: Wiley; 1965.
- Lambert M, Duppong Hurley K, Gross T, Epstein M, Stevens A. Validation of the Symptoms and Functioning Severity Scale in residential group care. *Administration and Policy in Mental Health and Mental Health Services Research*. 2014 Advance online publication. 10.1007/s10488-014-0575-z
- Larzelere RE, Daly DL, Davis JL, Chmelka MB, Handwerk ML. Outcome evaluation of Girls and Boy's Town Family Home Program. *Education and Treatment of Children*. 2004; 27:130–149.
- McKinney C, Morse M. Assessment of disruptive behavior disorders: Tools and recommendations. *Professional Psychology: Research and Practice*. 2012; 43:641.

- Muthén, B.; Satorra, A. Complex sample data in structural equation modeling. In: Marsden, PV., editor. *Sociological methodology*. Washington, DC: American Sociological Association; 1995. p. 267-316.
- Nunnally, JC. *Psychometric theory*. 2nd. New York, NY: McGraw-Hill; 1978.
- Overington L, Ionita G. Progress monitoring measures: A brief guide. *Canadian Psychology/ Psychologie Canadienne*. 2012; 53:82–92.
- Robst J, Armstrong M, Dollard N. Comparing outcomes for youth served in treatment foster care and treatment group care. *Journal of Child and Family Studies*. 2011; 20:696–705.
- Shaffer D, Fisher P, Lucas C, Dulcan M, Schwab-Stone M. NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39:28–38. [PubMed: 10638065]
- Straus MA, Douglas EM. A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence Victims*. 2004; 19:507–520. [PubMed: 15844722]
- Tarren-Sweeney, Michael. The Assessment Checklist for Children - ACC: A behavioral rating scale for children in foster, kinship and residential care. *Children and Youth Services Review*. 2007; 29:672–691.
- Tarren-Sweeney M. The Assessment Checklist for Adolescents—ACA: A scale for measuring the mental health of young people in foster, kinship, residential and adoptive care. *Children and Youth Services Review*. 2013a; 35:384–393.
- Tarren-Sweeney M. The Brief Assessment Checklists (BAC-C, BAC-A): Mental health screening measures for school-aged children and adolescents in foster, kinship, residential and adoptive care. *Children and Youth Services Review*. 2013b; 35:771–779.
- Turner W, Macdonald G. Treatment foster care for improving outcomes in children and young people: A systematic review. *Research on Social Work Practice*. 2011; 21:501–527.
- Wolf M, Kirigin K, Fixsen D, Blasé K, Braukmann C. The Teaching-Family Model: A case study in data-based program development and refinement (and dragon wrestling). *Journal of Organizational Behavior Management*. 1995; 15:11–68.
- Wolf M, Phillips E, Fixsen D, Beukmann C, Kirigin K, Wilner A, Schumaker J. Achievement place: The teaching-family model. *Child Care Quarterly*. 1976; 5:92–103.
- Wood, F.; Flowers, L.; Meyer, M.; Hill, D. How to evaluate and compare screening tests: Principles of science and good sense; 2002, November; Paper presented at the meeting of the International Dyslexia Association; Atlanta, GA.
- Volpe RJ, Gadow KD. Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review*. 2010; 39:350–363.

Table 1

Binary Classification Results

SFSS Form	n	Sensitivity	Specificity	Accuracy
Clinician				
Form A	141	0.89	0.89	0.89
Form B	141	0.80	0.97	0.92
Youth				
Form A	135	0.94	0.91	0.93
Form B	135	0.95	0.88	0.92

Note. Classification low-severity and elevated-severity (medium severity scores or higher).