

Psychometric Evaluation of a Pragmatic Measure for Assessing Adherence to System of Care Principles in Behavioral Health Service Interactions

Journal of Emotional and Behavioral Disorders
1–14

© Hammill Institute on Disabilities 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10634266211028204

jebd.sagepub.com



Nathaniel J. Williams, PhD¹ , James Beauchemin, PhD¹,
Guido Giuntini, MS¹, Jennifer Griffis, MLS², and Ya Mo, PhD¹

Abstract

Provider adherence to system of care principles in service interactions with families is an important indicator of behavioral health service quality for youth; however, valid and pragmatic measures suitable for monitoring this quality indicator at population scale have not been developed. This article reports on two studies that developed and evaluated such a measure. In Study 1, an iterative, family-partnered process resulted in generation of 18 items that demonstrated unidimensionality and strong reliability among caregivers of youth participating in behavioral health services ($N = 141$). In Study 2, data from a second, statewide, stratified random sample of caregivers ($N = 351$) confirmed the items' unidimensionality, discriminant validity, and criterion-related validity. Higher scores on the System of Care Adherence Scale were associated with lower risk of youth psychiatric hospitalization, greater perceived improvement in youth functioning, and greater increases in caregivers' self-efficacy to access services. Item response theory analyses indicated the items were strongly related to adherence; however, most were optimal for differentiating between low to moderate levels of adherence. The System of Care Adherence Scale is a psychometrically sound measure suitable for population surveillance of the extent to which families experience system of care principles in their interactions with providers.

Keywords

of care, system(s), implementation, programs/practices, adherence, psychometrics, involvement, families/parent(s)

Introduction

Communities in the United States have long struggled to develop service systems that effectively meet the needs of children and adolescents who experience emotional and behavioral disorders, particularly those 6%–10% of youth who experience the most severe impairment and the greatest risk of long-term negative outcomes (Knitzer, 1982; U.S. Department of Health and Human Services, 1999; Williams et al., 2018). Only 44%–53% of youth who are most affected by psychiatric disorders receive any behavioral health services within a given 12-month period (Merikangas et al., 2010) and recent data show that usage of highly restrictive forms of care, such as inpatient hospitalizations, increased by 18%–30% from 2009 to 2014 (Teich et al., 2018). There is also ample evidence of high variability in the quality and effectiveness of community behavioral health services for youth (Garland et al., 2013; Ogles et al., 2008). These deficits in care have prompted several national

organizations, including the U.S. Substance Abuse and Mental Health Services Administration, the National Quality Forum, the Institute of Medicine, and the U.S. Agency for Healthcare Research and Quality, to call for the development and scale-up of valid, reliable, and pragmatic measures to assess and monitor behavioral health service quality for youth at a population level (Patel et al., 2015).

An important indicator of service quality in children's behavioral health is the extent to which services are provided in accordance with system of care principles (Stroul & Friedman, 1986). A system of care is

¹Boise State University, Boise, ID, USA

²Raising Superkids LLC, Grangeville, ID, USA

Corresponding Author:

Nathaniel J. Williams, School of Social Work, Boise State University,
1910 W. University Drive, Boise, ID 83725, USA.

Email: natewilliams@boisestate.edu

a spectrum of effective, community-based services and supports for children and youth with or at risk for mental health or other challenges and their families, that is organized into a coordinated network, builds meaningful partnerships with families and youth, and addresses their cultural and linguistic needs, in order to help them to function better at home, in school, in the community, and throughout life. (Stroul et al., 2010)

Services delivered within systems of care are guided by principles believed to support positive youth outcomes, including: family-driven, child-centered, individualized, strengths-based, culturally and linguistically competent, community-based (i.e., provided in the least restrictive environment), accessible, and collaborative and coordinated (Pires, 2002). System of care principles are designed to be operationalized at the level of both policy and direct practice (Hernandez et al., 2001; Stroul & Friedman, 1986). At the policy level, principles guide how services are funded, organized, and managed. At the direct practice level, principles guide how care providers interact with youth and families and how interventions are framed and coordinated.

Adherence to system of care principles refers to the extent to which these principles are implemented as intended by the model developers (Kutash et al., 2011; Schoenwald et al., 2011). Because the principles can be operationalized at the levels of policy and direct practice, adherence can be assessed at both levels (Hernandez et al., 2001; Kutash et al., 2011). Preliminary evidence suggests there is considerable variation in adherence to system of care principles across geographic regions at both the policy (Greenbaum et al., 2011; Vinson et al., 2001) and direct practice levels (Hernandez et al., 2001). At the policy level, greater adherence to system of care principles is associated with better access to services for youth, greater continuity of care, and less restrictive care (Bickman, 1996). At the direct practice level, greater adherence to system of care principles in practitioner interactions with families is associated with superior improvements in youth symptoms and functioning (Hernandez et al., 2001; Stephens et al., 2004).

Although adherence to system of care principles is an important quality indicator for youth behavioral health services, there are few psychometrically validated tools that assess youths' and families' experiences of system of care principles at the point of interaction with service providers (Patel et al., 2015). Tools such as the System of Care Implementation Survey (Greenbaum et al., 2011) focus on adherence at the policy level by asking selected formal and informal leaders of child-serving systems and organizations about their perceptions of service funding, organization, and delivery (Kutash et al., 2011). Such tools offer insights into the views of system leaders; however, they may not be as useful for evaluating adherence at the level of service interaction. Other tools, such as the System of

Care Practice Review (Hernandez et al., 2001; Stephens et al., 2004), assess adherence at the point of service interaction, but use a time- and resource-intensive case review methodology that employs trained coders to conduct interviews and review documentation. This approach offers in-depth insights about specific family experiences but may be less useful as a population-level quality surveillance measure within a public health framework (Perou et al., 2013). Measures designed to assess adherence to system of care principles are distinct from other measures such as those evaluating satisfaction with services (e.g., Brannan et al., 1996), change in youth symptoms (e.g., Achenbach & Ruffle, 2000), or provider adherence to specific program models such as Wraparound (e.g., Bruns et al., 2004).

Study Aims

The goal of this research was to develop and psychometrically evaluate a measure of adherence to system of care principles at the level of direct practice that would be suitable for population surveillance of this quality indicator within a public health framework. Adherence was defined as the extent to which service interactions experienced by youth and families engendered system-of-care principles. Given the importance of family voice as a guiding principle within systems of care, and drawing on evidence from research showing that caregivers' ratings of treatment adherence predict variation in service outcomes for youth (Henggeler et al., 1997; Lange et al., 2019), the study aimed to develop a pragmatic, caregiver-reported measure. Alongside reliability and validity, pragmatism is an important feature of implementation measures if they are to be used for sustained quality monitoring and performance measurement (Stanick et al., 2018).

In Study 1, a team of researchers, system leaders, and families worked together to define the content domain to be assessed, generate and refine items, and evaluate the dimensionality and reliability of the items within a community sample of families receiving behavioral health services. In Study 2, the team refined the items and tested their structural, discriminant, and criterion-related validity in a second, independent statewide sample of caregivers of youth participating in services. In addition, item response theory was used in Study 2 to evaluate the performance of individual items and to select items for a short-form of the scale suitable for large-scale population assessment of system of care implementation.

Study 1

Method

Setting. Study 1 was completed in 2019 within a statewide system of care in a western State in the United States. The project was initiated and funded by State leaders with the

goals of (a) developing a valid, reliable, and pragmatic tool for assessing and monitoring the quality of system of care implementation at a statewide, population level; (b) creating a mechanism for the inclusion of families' voices in the evaluation of service delivery; and (c) generating targets for service improvement. Toward this end, a team of researchers ($n = 2$), program managers ($n = 2$), clinicians ($n = 2$), and family members ($n = 2$) was assembled to define criteria for adherence to system of care principles and generate a methodology for collecting data. Given the overarching goals, a pragmatic, quantitative survey was developed to be completed by caregivers of youth. Survey methodology was selected to facilitate the collection of population representative data on families' experiences of care to monitor changes in system performance over time.

Definition of system of care adherence criteria and item generation. The study team generated criteria for evaluating adherence to system of care principles at the point of service interactions through an iterative process that began with a review of policy documents, monographs, and articles describing system of care principles and their operationalization, including the target system's practice manual (Pires, 2002; Stroul & Friedman, 1986). In total, eight criteria were identified that addressed the extent to which families experienced services as (a) family-centered, (b) strengths-based, (c) youth-focused, (d) community based, (e) accessible, (f) collaborative and coordinated, (g) culturally competent, and (h) outcome oriented. Following operationalization of the criteria, literature was reviewed to identify potential items that assessed the target domains. In total, 181 potentially relevant items or item stems were identified from a variety of sources including measures assessing system of care implementation at the community level, system of care intensive case review protocols, and fidelity assessments for adjacent service models that reflect system of care principles (e.g., case management and Wraparound). Next, the fit of items with specified content domains was evaluated, item wording was revised to address direct practice and targeted criteria, and new items were generated to ensure adequate coverage. Items were evaluated for clarity and reading level, and wording was revised until the measure achieved a Flesch-Kincaid Grade level of 5.9. The final item pool included 18 items which were new or significantly modified from the original sources. As a final step, items were pilot tested with clinicians and families to evaluate their relevance for practice and interpretability before field testing.

Participants and procedure. Review and approval of the study procedures was provided by the Institutional Review Board (IRB) at the first author's institution. To evaluate reliability and dimensionality, items were administered via a postal mail survey to a statewide population of caregivers

of youth who participated in community behavioral health services delivered through the State's regional offices. The sampling frame ($N = 1,087$) included all households that participated in services from March 2018 to February 2019 and had a valid mailing address on file ($n = 5$ households did not have a valid mailing address).

The survey was fielded from April 2019 to May 2019 following guidelines described by Dillman et al. (2014). All households received a signed, pre-survey letter printed on State letterhead informing them that a survey would be arriving in 1 week and requesting participation. At subsequent 1-week intervals, households received mailings consisting of (a) a signed invitation letter, survey, and business reply envelope; (b) printed postcard reminder; and (c) for households that had not yet responded, a final follow-up letter with a new survey and business reply envelope. Caregivers were instructed to complete the survey for a single eligible youth; if caregivers had two or more youths in services, they were asked to complete the survey for the youth whose birthday occurred next in the calendar year.

Measures. *Adherence to system of care principles* was assessed using the 18 items developed for this study as described above (hereafter referred to as the System of Care Adherence Scale). Each item included a statement that described the extent to which the family experienced system of care principles in their direct interactions with their care provider (see Table 1 for items). Caregivers indicated the extent to which they disagreed or agreed with each statement on a 5-point Likert-type scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). All statements were affirmatively worded to minimize response burden.

Perceived improvement in youth functioning was measured using four items developed for this study and modeled after other population surveys of behavioral health system outcomes (e.g., the CAHPS Experiences of Care and Health Outcomes survey developed by the Agency for Health care Research and Quality, 2019). Items assessed caregiver perceptions of the extent to which the target youth's functioning had improved at home, at school, in the community, and in terms of their overall mental health during the last 6 months (see Table 1 for items). Each item was scored on a 5-point Likert-type scale from 1 ("Much Worse") to 5 ("Much Better").

Data analysis. Classical test theory and item response theory (IRT) both assume that the set of items constituting a scale measure a single underlying latent construct, referred to as unidimensionality (DeVellis, 2016; Hambleton et al., 1991). To evaluate the dimensionality of the 18 items assessing adherence to system of care principles and the 4 items assessing perceived improvement in youth functioning, principal components exploratory factor analysis was utilized (Bryant & Yarnold, 1995). The number of factors to extract was

Table 1. Item Loadings From Principal Components Exploratory Factor Analysis (Study 1) and Confirmatory Factor Analysis (Study 2).

Item no. Study 1	Item no. Study 2	Scale and Items	Study 1 (N = 141)		Study 2 (N = 351) ^a
			Factor 1	Factor 2	
System-of-Care Adherence Scale					
1	1	The goals we are working on with the provider are the ones I believe are most important for my child/youth.†	0.662	0.109	0.840
2	2	The provider encourages me to share what I know about my child/youth's strengths and needs.†	0.819	0.067	0.819
3	3	The services my child/youth receives focus on what he/she is good at, not just on problems.†*	0.736	0.163	0.731
4	4	The assessment completed by the provider accurately represents my child/youth's needs.†*	0.753	0.038	0.857
5	5	Meetings with the provider occur at times and locations that are convenient for me.†	0.727	0.095	0.763
6	6	My child/youth is an active participant in planning his/her services.†*	0.553	0.402	0.633
8	8	The provider makes sure everyone on my child's treatment team is working together in a coordinated way.†*	0.837	0.108	0.872
9	9	My child and I are the main decision-makers when it comes to planning my child/youth's services.†	0.706	0.261	0.757
10	10	Services we receive are respectful of our family's language, religion, race/ethnicity, and culture.†	0.772	-0.120	0.788
11	11	The provider often works with our family to measure my child/youth's progress toward his/her goals.†	0.817	0.175	0.897
12	12	When services are not helping, the provider leads my child/youth's team in a discussion of how to make things better.†*	0.785	0.283	0.852
13	13	The provider talks with us about how we can use things we are good at to overcome problems.†*	0.809	0.280	0.882
14	14	When decisions are made about services, my child/youth has the opportunity to share his/her own ideas.†*	0.670	0.373	0.764
15	15	The provider suggests changes in my child/youth's treatment plan or services when things aren't going well.†*	0.793	0.202	0.869
17	17	The provider makes specific suggestions about what services might benefit my child/youth.†	0.820	0.207	0.861
18	18	I know who to contact for help if I have a concern or complaint about my provider.†	0.611	0.076	0.697
7		In times of crisis, my youth's provider is one of the first people I (would) call.	0.691	0.151	
16		My child/youth has a safety plan that I believe will help us in times of crisis.	0.577	0.460	
	7	The provider respects me as an expert on my child/youth.†			0.812
	16	The provider demonstrates hope and optimism in meetings with my family.			0.906
	19	My family can easily access the services my child needs most.†*			0.761
	20	The provider who has been working with my child and family shows that he/she will not give up on us.			0.914
	21	I feel that the provider openly and honestly communicates with my family.			0.901
Perceived Improvement in Youth Functioning Scale					
21	31	Compared to 6 months ago, how would you rate your youth's behavior at home now?	0.081	0.858	0.885
22	32	Compared to 6 months ago, how would you rate your youth's performance at school now?	0.098	0.822	0.706
23	33	Compared to 6 months ago, how would you rate your youth's behavior in the community now?	0.164	0.831	0.811
24	34	Compared to 6 months ago, how would you rate your youth's overall mental health now?	0.114	0.895	0.910

Items marked with a (†) were included in the final 18-item long-form version of the System of Care Adherence Scale. Items marked with an (*) were included in the final 9-item short-form version of the scale.

^aAll loadings are statistically significant at $p < 0.001$.

determined using parallel analysis (Hayton et al., 2004; Horn, 1965). In addition, the scree plot and the magnitude of item loadings were examined (Kahn, 2006). Internal consistency reliability was evaluated using Cronbach's alpha. All analyses were conducted in SPSS version 25.

Results

In total, 141 caregivers returned surveys rating their service experiences. Excluding non-deliverable mail (13%), this represented a response rate of 15%. On average, youths reported on by caregivers were 14.4 years old ($SD = 2.8$ years) with a median of 13 months in services ($SD = 32.7$ months). Almost half the sample was female ($n = 68, 48\%$) and the race and ethnicity distribution of the sample reflected the State's youth population: 85% ($n = 121$) of youth were identified as White (less than 5% were identified in each of the other race categories) and 13% ($n = 18$) were identified as Hispanic/Latino.

Dimensionality. Results of the parallel analysis indicated two factors should be extracted and retained based on these factors achieving eigenvalues greater than those generated from the random data ($p < .05$). The scree test results also supported a two-factor solution; consequently, two factors were extracted and retained. Examination of the factor loadings from the two-factor varimax-rotated solution indicated that items on the two extracted factors coalesced around the two hypothesized constructs of (a) adherence to system of care principles and (b) perceived improvement in youth functioning. Standardized loadings of the adherence items were uniformly high on a single factor ($M = 0.73$, $min = 0.55$, $max = 0.84$) and low on the other factor ($M = 0.19$, $min = -0.12$, $max = 0.40$), with one exception. The item "My child/youth has a safety plan that I believe will help us in times of crisis," exhibited moderate loadings on both factors, resulting in elimination from the adherence scale. Standardized loadings of the four items assessing improvement in youth functioning also exhibited very high loadings on a single factor ($M = 0.85$, $min = 0.82$, $max = 0.90$) and low loadings on the other factor ($M = 0.11$, $min = 0.08$, $max = 0.16$). To confirm the results, the exploratory factor analysis was re-estimated excluding the omitted item and results were substantively identical (standardized factor loadings for adherence items = 0.56 to 0.84; loadings for perceived outcome items = 0.82 to 0.89). These results supported the unidimensionality of the items assessing adherence to system of care principles, and the unidimensionality of items assessing perceived improvement in youth functioning; they indicate the system of care adherence items are measuring a single, underlying latent construct.

Reliability. Excluding the safety planning item, coefficient alpha for the 17 system of care adherence items was excellent ($\alpha = .95$). Coefficient alpha for the four perceived

improvement in youth functioning items was also very good ($\alpha = .89$).

Study 2

In Study 2, the system of care adherence items were further refined and psychometric properties of scores on the items were evaluated within a second, independent, statewide sample of youth participating in publicly funded behavioral health services. It was hypothesized that higher System of Care Adherence Scale scores would be positively related to greater perceived improvement in youth functioning and caregiver self-efficacy to access services and negatively related to youth risk for psychiatric hospitalization or other out-of-home placement.

Method

Item modification. Following the promising results of Study 1, research team members reconvened to further evaluate the adequacy of the system of care adherence items relative to the system's needs for quality monitoring and to plan for a larger-scale psychometric evaluation. Drawing on the adherence criteria and definitions generated in Study 1, the team evaluated the extent to which each domain was sufficiently covered by the 17 items. One item was dropped for conceptual reasons ("In times of crisis, my youth's provider is one of the first people I [would] call") and five new items were generated to improve the measure's assessment of service accessibility ("My family can easily access the services my child needs most") as well as the nature of relationships between providers and families (e.g., "The provider demonstrates hope and optimism in meetings with my family"). In total, 21 system of care adherence items were included in Study 2 (see Table 1).

Participants and procedures. The target population for Study 2 included all youth and families within the State who had participated in system of care behavioral health services from July 1, 2019 to Dec 30, 2019. Within the State, eligibility for system of care services is determined through administration of the Child and Adolescent Needs and Strengths (CANS) assessment (Anderson et al., 2003), which is implemented across multiple child-serving systems (Medicaid behavioral health carve out, state Division of Behavioral Health, juvenile justice, child welfare). These CANS assessments are housed within a unified statewide database which formed the sampling frame for Study 2. The frame included a total of $N = 9,373$ youth who had received a CANS assessment and been deemed eligible to participate in system of care services during the study dates.

Sample size determination for Study 2 was based on guidelines for the psychometric evaluation of items within the item response theory literature (Nguyen et al., 2014).

Multiple authors have indicated that samples of at least 300 to 350 respondents are needed to generate stable parameter estimates (e.g., Edelen & Reeve, 2007). Based on these guidelines, and assuming an approximately 10% response rate, $N = 3,999$ youth were sampled. To ensure adequate representation across geographic regions, the study employed stratified random sampling with proportional allocation across the State's seven administrative regions.

Items for Study 2 were administered through a postal mail survey fielded from February 2020 to April 2020. Similar to Study 1, a signed, pre-survey letter, printed on state letterhead, was sent to potential participants 1 week before the survey, requesting caregivers' participation. One week later, caregivers received the survey with a signed invitation letter and a business reply envelope. Caregivers were asked to complete the survey for a specific eligible youth and to return their responses in a supplied business reply envelope. All study procedures were reviewed and approved by the affiliated IRB.

Measures

Adherence to system of care principles. For Study 2, the System of Care Adherence Scale included 16 of the original items developed in Study 1 as well as 5 new items generated as described above (see Table 1 for items). Given the variability in service providers across this heterogeneous sample, caregivers were asked to identify the service provider who had worked with their youth most frequently during the previous 6 months and to rate their experiences with that provider. To facilitate recall and accuracy, caregivers were asked to indicate the specific type of provider (e.g., counselor, medication prescriber) immediately prior to making their ratings. Item format and response options for the 21 items were identical to Study 1. Coefficient alpha for the 21 system of care adherence items included in Study 2 was excellent ($\alpha = .96$).

Perceived improvement in youth functioning. Caregivers reported on their youth's perceived improvement in daily functioning using the same four items developed in Study 1. For analysis, the mean of these items was calculated to assess overall perceived improvement in the youth's functioning during the previous 6 months. Coefficient alpha for these items was very good ($\alpha = .87$).

Caregiver self-efficacy to access services. Caregivers reported on change in their self-efficacy to access services and supports for their youth using a single item developed for this study, "Compared to 6 months ago, how would you rate your ability to effectively access services and supports your child/youth needs?" Response options ranged from 1 ("Much Worse") to 5 ("Much Better").

Youth out-of-home placements. Caregivers responded to two questions regarding out of home placements for their youth: (a) "In the last 6 months, how many total nights did your child/youth spend in a hospital due to problems with behaviors or feelings?" (None, 1 to 2, 3 to 7, 8 or more) and (b) "In the last 6 months, how many times has your child/youth had a new out-of-home placement (such as juvenile detention, psychiatric hospital, or treatment center) OR moved between out-of-home placements?" (0, 1, 2, 3 or more). Because few youths had experienced one or more psychiatric hospitalizations ($n = 20, 6\%$) or other new out-of-home placements ($n = 25, 7\%$), responses to each question were dichotomized into 0 (None) and 1 (one or more) for analysis.

Control variables. The following variables were included in analyses assessing criterion-related validity to test associations between scores on the System of Care Adherence Scale and youth outcomes after adjusting for potential confounders. Youth characteristics included: youth age in years (coded as 0 = "<11 years," and 1 = "11 years or more" for analysis), youth sex (coded as 0 = "male," and 1 = "female"), youth race (coded as 0 = "White," 1 = "youth of color" for analysis), length of participation in services (coded as 0 = "<18 months," 1 = "18 months or more"), and severity of current impairment (coded as 0 = "caregiver indicated youth did not need a crisis intervention plan," 1 = "caregiver indicated youth needed a crisis intervention plan"). Models also included the provider rated by caregivers coded into three groups: outpatient provider (i.e., therapist or prescriber), community-based provider (i.e., case manager, home-based services, more than one provider), or other/unknown.

Data analysis

Dimensionality. Confirmatory factor analysis (CFA) was used to evaluate the dimensionality of the 21 system of care adherence items and the 4 perceived improvement in youth functioning items. Based on the hypothesized two-factor structure, a correlated two-factor CFA model was specified, with system of care adherence items forced to load on one factor and perceived improvement in youth functioning items forced to load on a second factor. Because the item responses were ordinal, weighted least square mean and variance adjusted (WLSMV) estimation was used in Mplus Version 8.0 (Muthén & Muthén, 2017); this estimator is appropriate for categorical outcomes (Li, 2016). Model fit was evaluated using the root mean square error of approximation (RMSEA), comparative fit index (CFI), and the standardized root mean square residual (SRMR; Schreiber et al., 2006). For RMSEA, commonly accepted criteria indicate values <0.05 demonstrate close fit, values <0.08 demonstrate reasonable fit, and values >0.10 demonstrate

poor fit (Kline, 2015; Schreiber et al., 2006). For CFI, values above 0.95 are typically accepted as indicative of good fit and values of SRMR <0.05 are also typically accepted as indicative of good model fit (Schreiber et al., 2006). An alternative one-factor model was estimated to assess discriminant validity; if the one-factor model adequately fit the data, it would undermine the unidimensionality assumption for the system of care adherence items by suggesting that caregivers' responses to all of the items (i.e., adherence and perceived improvement) were caused by a single underlying construct (e.g., positive response bias).

Preliminary analyses indicated that relatively few respondents endorsed the *Strongly Disagree* or *Disagree* response options. Consequently, the appropriateness of combining *Strongly Disagree* and *Disagree* response options was assessed by comparing results of coefficient alpha, CFA, and item response theory analyses for five- versus four-category versions of the scale. Results were superior for the four-category version; therefore, the four-category version of the scale was analyzed for parsimony.

Item properties and short-form selection. Item response theory was used to evaluate the performance of individual system of care adherence items and select items for inclusion in a short form of the scale. Item response theory complements classical test theory (DeVellis, 2016) by providing tools to evaluate individual item difficulty and discrimination (Hambleton et al., 1991). In the present study, more difficult items require providers to deliver care that is more adherent to system of care principles in order for caregivers to express a higher level of agreement (Ostini & Nering, 2006). Clinically, some items may be less difficult because they represent a minimum standard of appropriate care, whereas other items may be more difficult because they represent advanced ways of incorporating system of care principles into service interactions. For polytomous items that have four response categories (i.e., *Strongly Disagree/Disagree*, *Neutral*, *Agree*, *Strongly Agree*), three difficulty parameters, called thresholds, are estimated for each item. Each threshold indicates the level of system of care adherence that must be experienced by the respondent to have a 0.5 probability of endorsing that level of agreement or higher. According to Hambleton and colleagues (1991), difficulty thresholds near -2 indicate an item is very "easy" (i.e., participants could experience a relatively low level of adherence to system of care principles and still agree), whereas thresholds of $+2$ indicate an item is very difficult (i.e., participants would have to experience a very high level of adherence to system of care principles in their interactions with the provider to agree with the item).

Item discrimination indicates how well the item differentiates between care that is adherent versus non-adherent to system of care principles at the item's specific level of difficulty. Higher discrimination indicates the item is more

strongly related to system of care adherence and better at differentiating between adherent versus non-adherent care. Baker (2001) described item discrimination values of <0.64 as low, 0.65 to 1.34 as moderate, 1.35 to 1.69 as high, and ≥ 1.70 as very high.

To determine the best IRT model for the data, the fit of two models was compared that are appropriate for items with a polytomous response format—the graded response model and the generalized partial credit model (Ostini & Nering, 2006). Selection of the final model was based on comparison of Bayesian Information Criterion (BIC) values (Edelen & Reeve, 2007; Kang et al., 2009). After selecting a model, the tenability of the unidimensionality and local dependence assumptions were evaluated by examining item-level fit statistics (i.e., Signed chi-square test) and item slopes (Nguyen et al., 2014; Orlando & Thissen, 2000). Statistically significant signed chi-square tests indicate the model is a poor fit for the item; slopes >4 indicate a potential violation of local dependence (Edelen & Reeve, 2007; Hambleton et al., 1991). In addition, category characteristic curves, item information functions, and the test information function were examined (Nguyen et al., 2014). Models were fit using marginal maximum likelihood (MML) estimation via an Expectation-Maximization (EM) algorithm (Bock & Aitkin, 1981) as implemented in the Multidimensional Item Response Theory (mirt) package in R (Chalmers, 2012).

Criterion-related validity evidence. The criterion-related validity of scores on the long- and short-form versions of the System of Care Adherence Scale were evaluated by using multiple regression and multiple logistic regression analyses. These analyses tested whether scores on the System of Care Adherence Scale predicted risk of youth psychiatric hospitalization, risk of any new youth out-of-home placement, perceived improvement in youth functioning, and improvement in caregiver self-efficacy to access services for their youth, within the last 6 months, as reported by caregivers and after controlling for potential confounders (i.e., youth characteristics listed above). For dichotomous outcomes, adjusted odds ratios are reported. For continuous outcomes, adjusted betas and the unique variance explained in the outcome by the System of Care Adherence Scale (i.e., change in R^2) are reported. Analyses were implemented in Mplus Version 8.0 which uses full information maximum likelihood estimation to handle missing data (Enders, 2010). Missing data were minimal on all control variables (0%–7%).

Results

A total of 352 caregivers returned surveys (9.4% response rate after excluding returned mail); however, one caregiver did not rate services, yielding an analytic sample of

$N = 351$. There were no significant differences in response rates across regions ($\chi^2 = 4.18, p = .690$, min = 8.0%, max = 11.4%). On average, youths reported on by caregivers were 11.2 years old ($SD = 3.7$ years) with a median of 18 months in services ($SD = 31.1$ months). Caregivers reported on youth who were female ($n = 154, 45\%$) and male ($n = 189, 55\%$) as well as youths from diverse racial and ethnic backgrounds reflective of the larger State, including youth identified as Hispanic/Latino ($n = 47, 14\%$) and non-Hispanic/Latino youth ($n = 289, 86\%$), as well as White/Caucasian ($n = 300, 85\%$), multiple races ($n = 26, 7\%$), Black/African American ($n = 6, 2\%$), other race ($n = 6, 2\%$), and unknown race/ prefer not to answer ($n = 14, 4\%$).

Dimensionality. Results from the CFA analyses supported the unidimensionality of the system of care adherence items. The hypothesized two-factor CFA model exhibited good fit to the data as indicated by all fit indices (RMSEA = 0.06, RMSEA 90% CI = 0.05–0.07, CFA = 0.98, SRMR = 0.04). All unstandardized factor loadings were statistically significant (p 's < .001) and the standardized factor loadings were high, ranging from 0.63 to 0.91 (see Table 1). The system of care adherence factor and the perceived improvement in youth functioning factor were correlated at $r = 0.38$, providing discriminant validity evidence indicating that the sets of items assessed different constructs. Further support for the hypothesized 2-factor model was provided by comparing it to a one-factor model in which all items were forced to load onto a single factor. The one-factor model did not fit the data well and was rejected based on all criteria (RMSEA = 0.15, CFA = 0.89, SRMR = 0.10).

IRT calibration. The calibration of the 21 system of care adherence items using the graded response model resulted in a superior BIC of 13,155.61 compared to a BIC of 13,343.52 for the generalized partial credit model; consequently, the graded response model was used to calibrate and evaluate the items. Examination of the item discrimination parameters (slopes) from the graded response model indicated that four items may have violated the local dependence assumption (i.e., slope values >4). Three of these were newly developed items for Study 2 and had slopes ≥ 4.5 ; consequently, they were eliminated and the remaining 18 items were recalibrated. Examination of the item-level Signed chi-square tests (Nguyen et al., 2014; Orlando & Thissen, 2000) from the 18-item calibration indicated that the model fit the remaining 18 items well after controlling for Type I error rates with the Benjamini–Hochberg adjustment (Benjamini & Hochberg, 1995). Furthermore, estimation of a new two-factor CFA model that included only the 18 items along with the 4 youth functioning items indicated that model fit was excellent (RMSEA = 0.06, RMSEA 90% CI = 0.06 to 0.07, CFI = 0.98, SRMR = 0.05) as was coefficient alpha for the 18 system of care

adherence items ($\alpha = .95$). Based on these results, the 18 system of care adherence items were accepted as the final set of items for the long form of the System of Care Adherence Scale (see Table 2).

Item properties. Figure 1 shows category characteristic curves for four sample items chosen to reflect the general pattern observed across all 18 items. The four lines in each plot indicate the probability of endorsing each of the four response options across different levels of adherence. As a respondent's experience of system of care adherence increases, the likelihood of endorsing each response option increases or decreases. Ideally, each response option has a distinct peak, and the peaks should be ordered from *Strongly Disagree/Disagree* to *Strongly Agree*. Examination of the 18 system of care adherence items indicated they fit this pattern well, supporting the use of four response options, all of which appear to provide meaningful information about respondents' experiences of care.

As is shown in Table 2, discrimination parameter estimates (i.e., slopes) for the 18 items ranged from 1.49 to 4.17 (corresponding to standardized factor loadings of 0.65 to 0.89). These values fell within the high to very high range and indicated that each item is (a) strongly related to the underlying construct of adherence to system of care principles and (b) discriminates well between care that is more versus less adherent at the item's level of difficulty. Although slope values ≈ 4 are high, these items were retained to maintain adequate coverage across content domains. As a robustness check, we refit the graded response model, excluding the items with very high discrimination parameter estimates, and the parameter estimates were substantively identical.

Difficulty thresholds for the response options for each of the 18 items reflected a sizable range of adherence to system of care principles (–2.34 to 0.87, see Table 2). However, as is shown in the right panel of the person-item map in Figure 2, 74% of the thresholds fell below a logit value of 0, indicating that these items are most effective at assessing low to moderate levels of adherence to system of care principles. The person-item map in Figure 2 also illustrates the frequency distribution of families' experiences of care in the left panel based on an IRT-scaled score with a mean of 0 and a standard deviation of 1. The plot indicates that families' experiences were normally distributed near the mean level of adherence with a small increase in the number of families at the top of the distribution ($n = 37, 10.5\%$) who experienced care that was highly adherent to system of care principles.

Short form selection. Items were selected for inclusion in the short-form of the System of Care Adherence Scale if their level of difficulty for the highest response option (i.e., *Strongly Agree*) was above the median. This resulted in

Table 2. Item Response Theory Parameter Estimates and Fit Statistics for System-of-Care Adherence Items.

Item no.	Item	<i>a</i>	<i>b</i> ₁	<i>b</i> ₂	<i>b</i> ₃	<i>S</i> - χ^2	<i>p</i> value
6 ^a	My child/youth is an active participant in planning his/her services.	1.49	-1.28	-0.35	0.87	55.61	0.791
19 ^a	My family can easily access the services my child/youth needs most.	2.28	-1.04	-0.45	0.52	70.08	0.039
14 ^a	When decisions are made about services, my child/youth has the opportunity to share his/her own ideas.	2.21	-1.50	-0.76	0.40	46.45	0.617
18	I know who to contact for help if I have a concern or complaint about my provider.	1.80	-0.95	-0.51	0.39	61.99	0.548
12 ^a	When services are not helping, the provider leads my child/youth's team in a discussion of how to make things better.	3.13	-1.23	-0.38	0.38	49.75	0.255
13 ^a	The provider talks with us about how we can use things we are good at to overcome problems.	3.63	-1.14	-0.58	0.33	34.18	0.799
15 ^a	The provider suggests changes in my child/youth's treatment plan or services when things aren't going well.	3.46	-1.20	-0.55	0.33	44.73	0.280
8 ^a	The provider makes sure everyone on my child's treatment team is working together in a coordinated way.	3.67	-1.24	-0.47	0.30	48.95	0.132
3 ^a	The services my child/youth receives focus on what he/she is good at, not just on problems.	2.02	-1.95	-1.02	0.27	26.83	0.986
4 ^a	The assessment completed by the provider accurately represents my child/youth's needs.	3.21	-1.57	-0.90	0.25	35.80	0.616
17	The provider makes specific suggestions about what services might benefit my child/youth.	3.28	-1.32	-0.78	0.21	61.65	0.026
11	The provider often works with our family to measure my child/youth's progress toward his/her goals.	4.17	-1.17	-0.64	0.18	34.32	0.548
9	My child and I are the main decision-makers when it comes to planning my child/youth's services.	2.26	-1.66	-0.99	0.12	56.02	0.292
5	Meetings with the provider occur at times and locations that are convenient for me.	2.27	-1.68	-0.96	0.01	61.22	0.155
7	The provider respects me as an expert on my child/youth.	2.69	-1.76	-1.07	-0.03	40.46	0.406
1	The goals we are working on with the provider are the ones I believe are most important for my child/youth.	3.02	-1.90	-1.31	-0.08	22.80	0.786
2	The provider encourages me to share what I know about my child/youth's strengths and needs.	2.69	-1.84	-1.16	-0.22	45.49	0.160
10	Services we receive are respectful of our family's language, religion, race/ethnicity, and culture.	2.49	-2.34	-1.54	-0.51	43.6	0.066

Note. *N* = 351. Parameters estimated using a graded response model. *a* = item slope/discrimination parameter; *b*₁ - *b*₃ = item difficulty thresholds; *S* - χ^2 = generalized signed Chi-square test (all *p* values were nonsignificant at *p* > .05 after correcting for multiple tests using the Benjamini-Hochberg false discovery procedure).

^aItem included in the short-form version of the scale.

selection of the nine most difficult items from the set (see Table 2) excluding Item 18 (“I know who to contact for help if I have a concern or complaint about my provider”). Item 18 was omitted in favor of including Item 4 (“The assessment completed by the provider accurately represents my child/youth's needs”) given that the latter was viewed as more informative for guiding improvements in clinical care. As evidenced by the standard error functions in Figure 3, the short form maintained similar precision of measurement to the long form across levels of adherence to system of care principles. Both scales yield more precise measurement near the average and lower values of adherence to system of care principles, and less precise measurement at higher values of adherence.

Criterion-related validity evidence. Results of the regression analyses assessing the relationships between the short and long forms of the System of Care Adherence Scale and the four outcome variables supported the criterion-related validity of scores on both scales. As expected, higher scores on the long-form System of Care Adherence Scale were positively related to greater perceived improvement in youth functioning ($B = 0.35, p < .001$), accounting for 9% of the variance above and beyond the control variables (full model $R^2 = 0.15$). In addition, higher scores on the long form were related to increased caregiver self-efficacy to access services for their youth ($B = 0.63, p < .001$), accounting for 19% of the variance beyond the control variables (full model $R^2 = 0.22$). Higher scores on the long

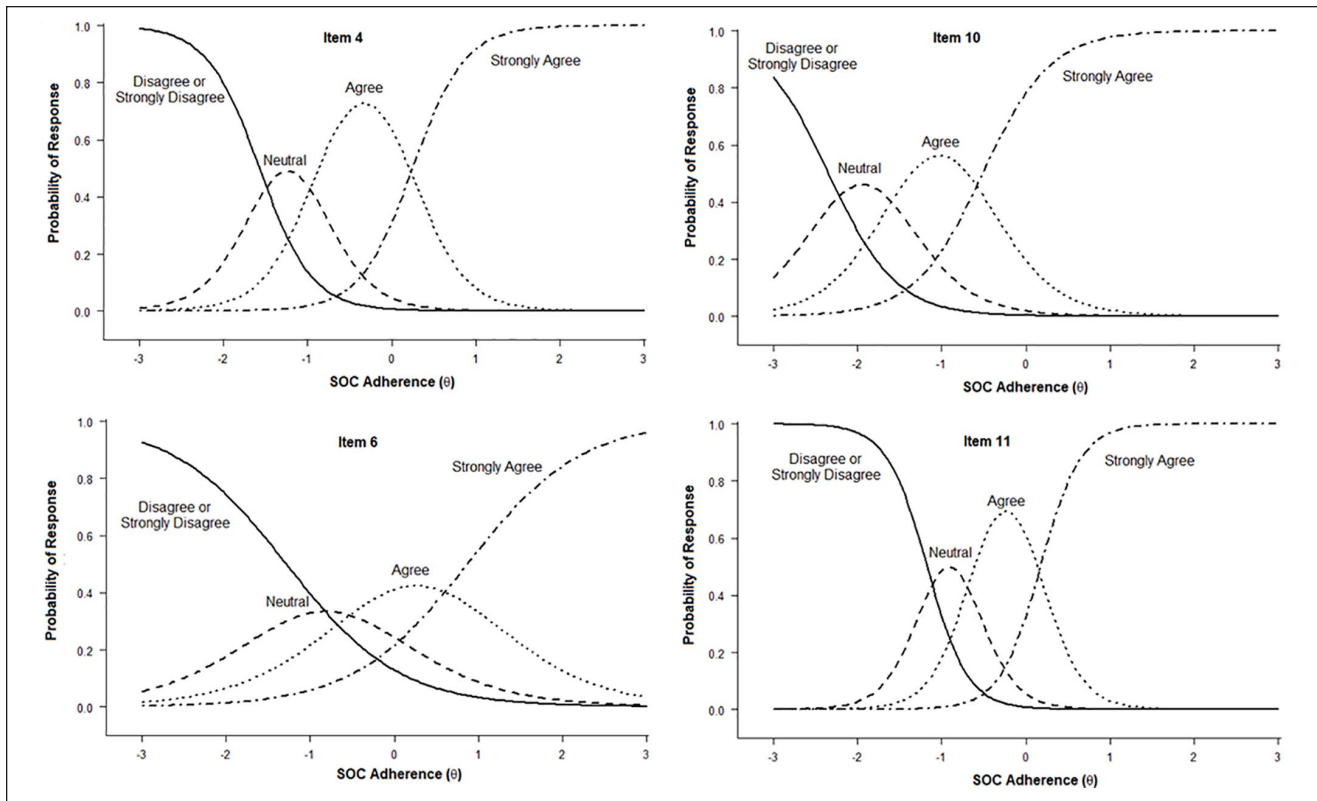


Figure 1. Category characteristic curves for four sample system of care (SOC) adherence scale items ($N = 351$).

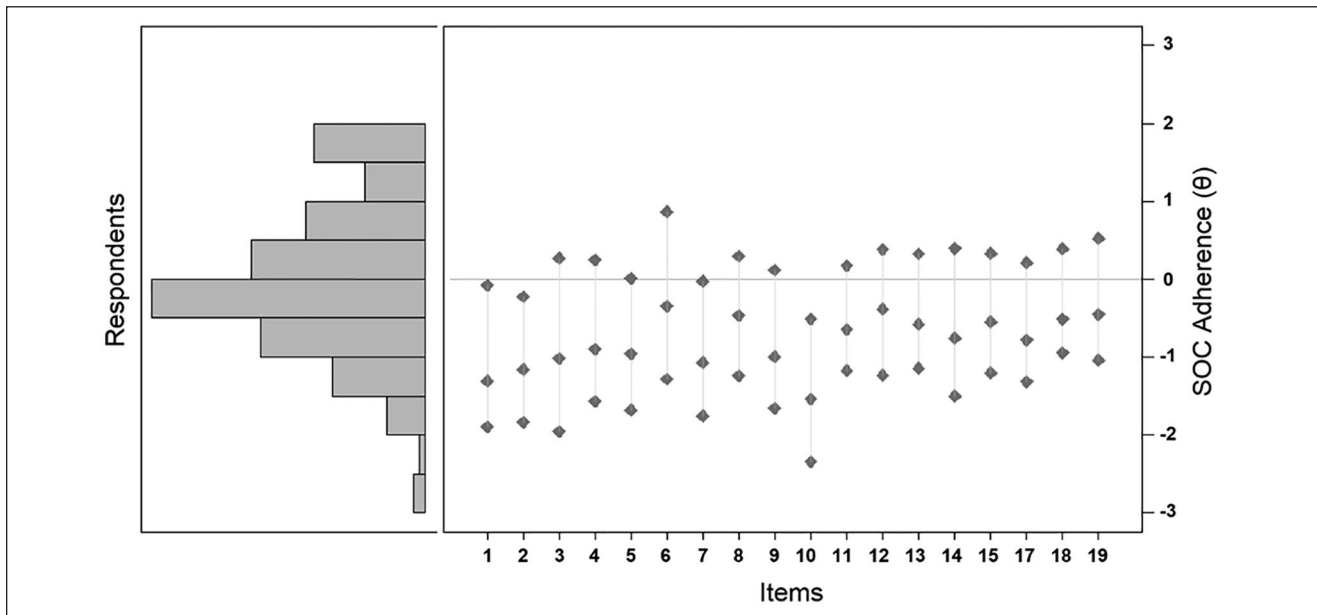


Figure 2. Person-item map for the 18-item System of Care (SOC) Adherence Scale ($N = 351$). Note. SOC adherence is measured in the logit scale.

form were also associated with *lower* odds of having experienced a new out-of-home placement within the last 6 months ($OR = 0.60, p = .049$); although the relationship

between the long form and risk of psychiatric hospitalization was not statistically significant ($OR = 0.61, p = .081$). On average, every one-point increase in system of care

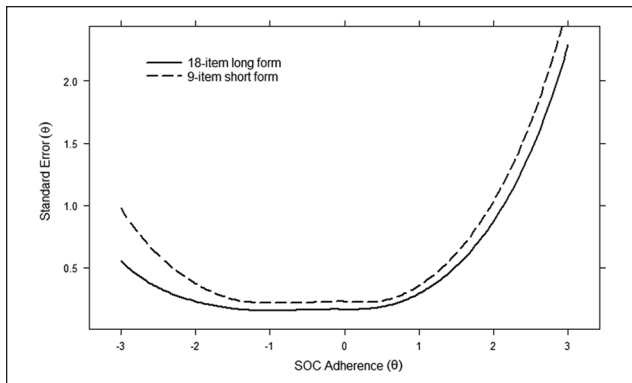


Figure 3. Standard error of measurement functions for the System of Care Adherence Scale ($N = 351$).

adherence experienced by the family (as measured by the long-form) was associated with a 40% decrease in the youths' odds of having experienced a new out-of-home placement within the last 6 months.

Scores on the short-form System of Care Adherence Scale were significantly related to all criterion variables in the hypothesized directions. Higher scores on the short-form predicted higher levels of perceived improvement in youth functioning ($B = 0.35, p < .001$), accounting for 11% of the variance (full model $R^2 = 0.17$). Higher scores on the short-form also predicted greater increases in caregivers' self-efficacy to access services ($B = 0.59, p < .001$), accounting for 19% of the variance (full model $R^2 = 0.22$). Youth who received care that was rated as more adherent to system of care principles on the short-form were significantly less likely to have experienced a psychiatric hospitalization ($OR = 0.58, p = .036$) or new out-of-home placement ($OR = 0.56, p = .015$) within the last 6 months. On average, for every one-point increase in system of care adherence on the short form, youths' odds of having experienced a psychiatric hospitalization decreased by 42%, and their odds of having experienced any new out-of-home placement decreased by 44%. These results provide criterion-related evidence supporting the validity of scores on the System of Care Adherence Scale.

Discussion

The studies reported in this article were designed to develop and evaluate a pragmatic measure of adherence of practice to system of care principles suitable for population surveillance of behavioral health service quality from the perspective of families. Based on system of care literature, a practice manual, and feedback from families, clinicians, and program administrators, items were generated to capture the extent to which families' interactions with behavioral health providers embodied system of care principles. Results of the exploratory factor analysis

from Study 1 supported the reliability and unidimensionality of the proposed system of care adherence items, indicating they tap a common underlying construct that is distinct from caregivers' perceptions of improvement in youth functioning. These findings were confirmed in Study 2 using confirmatory factor analysis of data from a second, independent, statewide sample of caregivers of youth receiving community behavioral health services. Results from Study 2 also provided criterion-related evidence supporting the validity of scores on the short- and long- forms of the System of Care Adherence Scale for predicting perceived improvement in youth functioning and increased caregiver self-efficacy to access services for their youth. Overall, results support the System of Care Adherence Scale as a psychometrically sound instrument for evaluating the extent to which direct service interactions experienced by youth and families engendered system-of-care principles. The System of Care Adherence Scale and scoring instructions can be found in Additional File 1 or may be obtained from NJW.

Results of the IRT analyses in Study 2 provided additional evidence supporting the psychometric properties of the System of Care Adherence Scale, and highlighted directions for continued scale development. Examination of the category characteristic curves for each of the 18 items confirmed that a modified four-category response format provided meaningful information about respondents' experiences of care. Accordingly, future iterations of the survey will include four response categories (*Disagree, Neutral, Agree, and Strongly Agree*) instead of five. In addition, examination of the difficulty thresholds for the 18 items indicated that these items are most effective in distinguishing between experiences of care at low to moderate levels of adherence to system of care principles (i.e., most items are easy to moderate in difficulty). Based on these results, efforts will be made to develop more difficult items; that is, items that require providers to deliver services that are highly adherent to system of care principles in order for caregivers to strongly agree with the item. Ideally, such items will allow the scale to distinguish between care that is good versus care that is truly exceptional.

An important attribute of any adherence measure is the ability to predict client outcomes (Schoenwald et al., 2011). In Study 2, scores on the System of Care Adherence Scale explained 9%–11% of the variance in youth's improvement in daily functioning during the last 6 months, a finding consistent with research from intensive case review methods (Stephens et al., 2004). These findings extend a growing number of studies which show that caregiver ratings of adherence to program models predict the extent to which youth benefit from services (Henggeler et al., 1997; Lange et al., 2019). In addition, higher scores on the long- and short-form System of Care Adherence

Scale predicted decreased odds of any youth out-of-home placement during the previous 6 months and higher scores on the short form predicted decreased odds of youth psychiatric hospitalization (42%). Previous research has shown that implementation of system of care principles at the community level results in lower usage of restrictive forms of care such as out-of-home placements (Bickman, 1996). However, these findings offer the first evidence linking increased adherence to system of care principles at the direct practice level with reduced likelihood of out-of-home placement. This is important given the emphasis within systems of care on reducing unnecessary, overly restrictive, and costly out-of-home placements (Pires, 2002).

The development and availability of a pragmatic measure of adherence to system of care principles at the point of service interactions is particularly timely given recent national initiatives calling for the development and validation of measures suitable for population surveillance of behavioral health service quality (Patel et al., 2015). Development of the System of Care Adherence Scale also fills a need for measures that evaluate the implementation of systems of care at the direct practice level from the perspective of families. The inclusion of families in the development and evaluation of the System of Care Adherence Scale and its focus on amplifying families' voices in service evaluation are important strengths of this measure (Stroul et al., 2010). Among the potential uses of the scale are providing feedback to providers who are learning the system of care principles and enabling stakeholders, including program site administrators, system leaders, and advocates, to assess system of care implementation across geographic regions. Such evaluations could inform system development and training as well as track improvement (or deterioration) in system of care implementation at the level of service interactions over time via longitudinal administration.

Use of the System of Care Adherence Scale should comprise only one component of a comprehensive multi-method assessment of adherence to system of care principles at both the policy and direct practice levels. Although research demonstrates that caregiver reports of adherence to program models predict service effectiveness (Henggeler et al., 1997; Lange et al., 2019), these reports nonetheless represent only one perspective and may be susceptible to erroneous recall or social desirability biases (Schoenwald et al., 2011). Use of multiple methods and triangulation can increase confidence in the findings of an evaluation. For example, fielding a population representative survey of families using the System of Care Adherence Scale and supplementing this information with additional data from targeted intensive case reviews (e.g., System of Care Practice Review, Hernandez et al., 2001) or policy-level

assessments of implementation (e.g., System of Care Implementation Survey; Greenbaum et al., 2011), would allow system leaders to robustly evaluate system of care implementation.

The results of this research should be interpreted within the context of the studies' limitations. Although the two studies reported here incorporated separate statewide samples of caregivers, the response rates for both mailed surveys were low (i.e., 15% and 9%, respectively). In theory, response bias from selective responding should not affect the psychometric properties of the System of Care Adherence Scale items as estimated via IRT. However, replication of these results with larger samples and with higher response rates will further strengthen confidence in the psychometric properties of the scale. To strengthen evidence of criterion-related validity, future research should test the associations between scores on the System of Care Adherence Scale and other validated system of care implementation measures such as the intensive System of Care Practice Review tool (Hernandez et al., 2001). Studies should also test whether System of Care Adherence Scale scores predict objective measures of youth outcome such as independently rated improvements in youth functioning or out-of-home service use verified by claims data; this was not possible in the present study because CANS data were only available at the point of service entry. Future research should also directly assess whether youths are placed out of home at the time of item completion to disentangle the potential confound between the types of services available in residential settings versus community settings and scores on the System of Care Adherence Scale. In addition, studies are needed to test for differential item functioning across youth and family characteristics. Although further research is needed, findings from these studies indicate that scores on the System of Care Adherence Scale provide a valid and reliable basis for making inferences about the extent to which families experienced system of care principles in their behavioral health care.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported in this publication was supported in part by the Idaho Department of Health and Welfare, Division of Behavioral Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Idaho Department of Health and Welfare.

ORCID iD

Nathaniel J. Williams  <https://orcid.org/0000-0002-3948-7480>

References

- Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review, 21*(8), 265–271.
- Agency for Healthcare Research and Quality. (2019). *CAHPS mental health care surveys*. <https://www.ahrq.gov/cahps/surveys-guidance/echo/index.htm>
- Anderson, R. L., Lyons, J. S., Giles, D. M., Price, J. A., & Estle, G. (2003). Reliability of the Child and Adolescent Needs and Strengths-Mental Health (CANS-MH) Scale. *Journal of Child and Family Studies, 12*(3), 279–289.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.
- Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist, 51*(7), 689–701.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459.
- Brannan, A. M., Sonnichsen, S. E., & Heflinger, C. A. (1996). Measuring satisfaction with children's mental health services: Validity and reliability of the satisfaction scales. *Evaluation and Program Planning, 19*(2), 131–141.
- Bruns, E. J., Burchard, J. D., Suter, J. C., Leverentz-Brady, K., & Force, M. M. (2004). Assessing fidelity to a community-based treatment for youth: The Wraparound Fidelity Index. *Journal of Emotional and Behavioral Disorders, 12*(2), 79–89.
- Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). American Psychological Association.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29.
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). SAGE.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(1), 5–18.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Garland, A. F., Haine-Schlagel, R., Brookman-Frazee, L., Baker-Ericzen, M., Trask, E., & Fawley-King, K. (2013). Improving community-based mental health care for children: Translating knowledge into action. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(1), 6–22.
- Greenbaum, P. E., Wang, W., Boothroyd, R., Kutash, K., & Friedman, R. M. (2011). Multilevel confirmatory factor analysis of the systems of care implementation survey (SOCIS). *Journal of Behavioral Health Services & Research, 38*(3), 303–326.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. SAGE.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. J. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821–833.
- Hernandez, M., Gomez, A., Lipien, L., Greenbaum, P. E., Armstrong, K. H., & Gonzalez, P. (2001). Use of the system-of-care practice review in the national evaluation: Evaluating the fidelity of practice to system-of-care principles. *Journal of Emotional and Behavioral Disorders, 9*(1), 43–52.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.
- Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, 34*(5), 684–718.
- Kang, T., Cohen, A. S., & Sung, H. J. (2009). Model selection indices for polytomous items. *Applied Psychological Measurement, 33*(7), 499–518.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.
- Knitzer, J. (1982). *Unclaimed children: The failure of public responsibility to children and adolescents in need of mental health services*. Children's Defense Fund.
- Kutash, K., Greenbaum, P. E., Wang, W., Boothroyd, R. A., & Friedman, R. M. (2011). Levels of system of care implementation: A national benchmark study. *Journal of Behavioral Health Services & Research, 38*(3), 342–357.
- Lange, A. M. C., van der Rijken, R. E. A., Delsing, M. J. M. H., Busschbach, J. J. V., & Scholte, R. H. J. (2019). Development of therapist adherence in relation to treatment outcomes of adolescents with behavioral problems. *Journal of Clinical Child and Adolescent Psychology, 48*(Suppl. 1), S337–S346.
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949.
- Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics, 125*, 75–81.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (Version 8).
- Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient-Patient-Centered Outcomes Research, 7*(1), 23–35.
- Ogles, B. M., Carlson, B., Hatfield, D., & Karpenko, V. (2008). Models of case mix adjustment for Ohio mental health consumer outcomes among children and adolescents.

- Administration and Policy in Mental Health and Mental Health Services Research*, 35(4), 295–304.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models* (No. 144). SAGE.
- Patel, M. M., Brown, J. D., Croake, S., Lewis, R., Liu, J., Patton, L., Potter, D. E. B., & Scholle, S. H. (2015). The current state of behavioral health quality measures: Where are the gaps? *Psychiatric Services*, 66(8), 865–871.
- Perou, R., Bitsko, R. H., Blumberg, S. J., Pastor, P., Ghandour, R. M., Gfroerer, J. C., & Huang, L. N. (2013). Mental health surveillance among children—United States, 2005–2011. *MMWR Supplements*, 62(2), 1–35.
- Pires, S. A. (2002). *Building systems of care: A primer*. National Technical Assistance Center for Children’s Mental Health, Georgetown University Child Development Center.
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health*, 38, 32–43.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338.
- Stanick, C. F., Halko, H. M., Dorsey, C. N., Weiner, B. J., Powell, B. J., Palinkas, L. A., & Lewis, C. C. (2018). Operationalizing the “pragmatic” measures construct using a stakeholder feedback and a multi-method approach. *BMC Health Services Research*, 18(1), Article 882.
- Stephens, R. L., Holden, E. W., & Hernandez, M. (2004). System-of-care practice review scores as predictors of behavioral symptomatology and functional impairment. *Journal of Child and Family Studies*, 13(2), 179–191.
- Stroul, B. A., Blau, G., & Friedman, R. M. (2010). *Updating the system of care concept and philosophy*. Georgetown University Center for Child and Human Development, National Technical Assistance Center for Children’s Mental Health.
- Stroul, B. A., & Friedman, R. M. (1986). *A system of care for severely emotionally disturbed youth*. Georgetown University.
- Teich, J. L., Mutter, R., Gibbons, B., Miller, K., Lynch, S., & Ali, M. M. (2018). Use of inpatient behavioral health services by children and adolescents with private insurance coverage. *Psychiatric Services*, 69(9), 1036–1039.
- U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General*. U.S. Department of Health and Human Services, U.S. Public Health Service.
- Vinson, N. B., Brannan, A. M., Baughman, L. N., Wilce, M., & Gawron, T. (2001). The system-of-care model: Implementation in twenty-seven communities. *Journal of Emotional and Behavioral Disorders*, 9(1), 30–42.
- Williams, N. J., Scott, L., & Aarons, G. A. (2018). Prevalence of serious emotional disturbance among US children: A meta-analysis. *Psychiatric Services*, 69(1), 32–40.