

The Factor Structure of the Ohio Scales: A Practical Measure of Psychological Symptoms in Youth

Francis Tony Bonadio¹ · Carolyn Tompsett¹

Published online: 17 September 2016
© Springer Science+Business Media New York 2016

Abstract Stakeholders have increased the focus on accountability for community mental health agencies, leading to the need for an efficient method of measuring mental health outcomes for youth. The Ohio Scales were developed specifically to address the drawbacks of existing measures of child and adolescent mental health such as length, cost, and ease of scoring. The current study examined the factor structure of the Ohio Scales Problem Severity scale and explored the concurrent validity of those factors by comparing them to similar factors on the Child Behavior Checklist and Youth Self Report, two well-established measures of childhood mental health. Using archival data from a community mental health agency serving youth and families, a randomly split sample was used to conduct an exploratory factor analysis on one subsample followed by a confirmatory factor analysis on a second subsample. The results of these analyses indicated that a four-factor model provided the best fit to the data for both youth and parent report on the Ohio Scales. The four factors included: Aggression, Delinquency, Depression, and Anxiety. Comparing the Ohio Scale factors to the factors on the Child Behavior Checklist and YSR revealed that theoretically similar factors (e.g., Anxiety and Internalizing) were more highly correlated than dissimilar factors (e.g., Anxiety and Externalizing). These results provide concurrent validity of the four-factor model of the Ohio Scales. These factors may provide more specific information

regarding the types of emotional and behavioral problems youth and parents are reporting to clinicians.

Keywords Psychometrics · Assessment · Child and adolescent mental health · Community mental health · Factor analysis

Introduction

Recent changes to the public health care system have created a focus on accountability for community mental health organizations (CMHs; Trask and Garland 2011). As resources dwindle, stakeholders increasingly demand measures of treatment efficacy in order for CMHs to retain funding (Dowell and Ogles 2008). However, implementing outcome measures for treatment comes with a cost in both expenses associated with purchase of measures and in clinician time dedicated to scoring and recording measure results (Trask and Garland 2011). The growing demand for tracking outcomes can be particularly challenging for over-utilized and under-staffed CMHs and other child and family serving agencies (Garland et al. 2003; Hatfield and Ogles 2004). Thus, there is a need for an empirically sound and practical measure of clinical outcomes. With this in mind, Ogles et al. (2000) developed the Ohio Scales as a practical yet empirically valid measure of children's problem severity and functioning.

Although many other measures of childhood emotional and behavioral problems are available, the Ohio Scales were designed specifically to meet the needs of CMHs. Ogles et al. (2000) collaborated with clinicians, clients, and stakeholders to design a measure which replicated the

✉ Francis Tony Bonadio
fbonadi@bgsu.edu

¹ Department of Psychology, Bowling Green State University, Bowling Green, OH, USA

strengths of other measures while addressing many of the drawbacks which limit the utility of these same measures in CMH settings. For example, the parent-reported Achenbach Child Behavior Checklist (CBCL; Achenbach 1991) is one of the most commonly used measures in clinical practice (Hatfield and Ogles 2004), and has many strengths including well-established psychometric properties, parallel forms for child self-report and teacher report, and the broad range of symptoms assessed (Achenbach and Rescorla 2004). Despite these strengths, the CBCL has many drawbacks that make it less functional in a CMH setting, including the length of the assessment (113 items), the time needed to complete the assessment (15–20 min for parents), the complex scoring of the measure, and the cost (Garland et al. 2003). The time needed to administer, score, and interpret the CBCL detracts from time spent on therapeutic intervention, and often, clinicians are not reimbursed for time spent scoring and interpreting measures. Another drawback of the CBCL can be the cost to CMHs, as forms, software, and licenses for scoring must be purchased regularly. Thus, although the CBCL provides an extensive measure of childhood mental health, these drawbacks limit the practical utility of the CBCL in a CMH setting (Garland et al. 2003).

The Ohio Scales were developed to address these drawbacks and to provide CMHs and other child and family serving agencies with an empirically sound measure that could be utilized for tracking outcomes throughout treatment. The Ohio Scales replicate the strengths of other measures—strong psychometric properties and multiple informants—while simultaneously addressing drawbacks such as considerations of time, expense, and ease of scoring. The result is a 48-item measure with three parallel forms (self-report, parent-report, and worker-report) that contain four independent scales: Problem Severity, Functioning, Client Hopefulness, and Satisfaction with Service (Ogles et al. 2000). The Problem Severity scale provides information on the frequency of common emotional and behavioral problems. The Functioning scale provides information on how well a youth is functioning within several areas of daily life (e.g., relationships, school). The Ohio Scales also include two subscales, Client Hopefulness and Satisfaction with Service, reflecting domains that tend to be less commonly incorporated into other outcomes measures but are of particular interest in clinical settings, as service providers are interested in monitoring service quality and client participation in services (Garland et al. 2007; Turchik et al. 2010). In order to address barriers such as cost and accessibility, the developers made the Ohio Scales available online and permit unlimited use for a minimal licensing fee. Many states' public mental health systems have adopted the Ohio Scales, possibly because they are tailored for CMHs (Turchik et al. 2007). Further,

since their development the use of the Ohio Scales has not been limited to traditional CMHs, but rather they have been utilized by other agencies which serve children and family including monitoring the effectiveness of wrap-around services (Bickman et al. 2003) and foster care agencies (Hayek et al. 2014).

The Problem Severity scale from the Ohio Scales is comparable to established measures of childhood mental health and problem behaviors (i.e., CBCL; Warnick et al. 2009), but it stands out due to its brevity and ease of scoring. When developing the items for the Problem Severity scale, four sources of information were considered: problem behaviors from the *Diagnostic and Statistical Manual, Fourth Edition, Text Revision* (DSM-IV TR; American Psychiatric Association 2000) used in diagnoses, a list of common presenting problems, items from other commonly-used assessments, and consultation with child service providers (Ogles et al. 2001). From these sources, the developers created 44 items, which were later reduced to 20 items for a short form of the Ohio Scales based on a request from service providers and a subsequent factor analysis. These items are based on common diagnostic symptoms and problem behaviors (e.g., “Feeling sad or depressed” and “Fits of anger”) which are rated based on frequency on a 6-point scale from 0 (*Not at All*) to 5 (*All of the Time*). The total Problem Severity score can easily be calculated by summing these twenty items into a total score ranging from 0–100, with higher scores indicating greater levels of problem severity. The simple scoring system requires no additional training, which is practical for CMH agencies as more paraprofessionals enter the field (Ogles et al. 2000). In addition to creating an easy scoring structure, a clinical cutoff was also established (i.e., above 25) as well as significant change scores (i.e., change of 10 or more) that can aid in the interpretation of the Ohio Scales (Ogles et al. 2000; Texas Department of Mental Health and Mental Retardation [TDMHMR] 2004).

Previous research has established the Ohio Scales Problem Severity scale as a reliable and valid measure of childhood mental health and problem behaviors (Dowell and Ogles 2008; Ogles et al. 2001; Turchik et al. 2007). Additionally, the Ohio Scales have also been compared to both the CBCL and the Strengths and Difficulty Questionnaire (SDQ; Ogles et al. 2000; TDMHMR 2004; Warnick et al. 2009). Ogles et al. (2000) found that in a small sample of parent reports ($N=28$) the Problem Severity scale from the Ohio Scales was highly correlated ($r=.89$) with the total score on the CBCL. In one study, when compared to the CBCL and SDQ (Warnick et al. 2009), the Ohio Scales Problem Severity scale had the highest specificity (88 %) in identifying youth who received a mental health diagnosis based on the Diagnostic Interview Schedule for Children, Version IV (DISC-IV). Further, the

Ohio Scales Problem Severity scale was the second best predictor of Disruptive Behavior diagnoses behind the CBCL Externalizing scale.

These findings support that the Ohio Scales are at least as effective as other assessments of childhood problem behaviors and mental health. However, unlike the other measures to which it was compared (e.g., CBCL, SDQ), the Ohio Scales Problem Severity scale does not have well-established subscales. Subscales such as Internalizing and Externalizing have been shown to add utility for treatment planning and for examining more specific outcomes (Turchik et al. 2007). In the previously mentioned study comparing the predictive utility of the CBCL, SDQ, and Ohio Scales, the Internalizing and Externalizing subscales on both the CBCL and SDQ were shown to be more predictive of diagnoses within their specific content areas than the overall score on the Ohio Scales Problem Severity scale (Warnick et al. 2009), suggesting that valid subscales would enhance the clinical utility of the Ohio Scales Problem Severity scale. However, only two non-peer-reviewed studies have begun to examine the factor structure of the Ohio Scales Problem Severity scale (Ogles et al. 2000; TDMHMR 2004). Analyses of the long form and the short form revealed a three-factor solution including Internalizing, Externalizing, and Conduct Disturbance/Delinquent Behaviors (Ogles et al. 2000; TDMHMR 2004). The factor analysis of the 20-item short form (TDMHMR 2004) demonstrated that only three items loaded on the Delinquent Behavior factor (i.e., “using drugs or alcohol”; “breaking rules or breaking the law”; “skipping school”). However, the CBCL has three similar items (i.e., “Drinks alcohol”; “Breaks rules”; “Truant”) that load on to the Externalizing factor (Achenbach and Rescorla 2004). Due to the small number of items on the third factor and factor loadings of similar items on other measures, we expected that a two-factor model might provide a better fit to the Ohio Scales Problem Severity scale. To the best of our knowledge, there have been no peer-reviewed publications replicating this factor structure and few studies exploring their validity. In one empirical study which used the subscales, the Delinquency, Externalizing, and Internalizing subscales discriminated among substance abuse, ADHD and disruptive behavior, and mood/anxiety disorders, respectively (Turchik et al. 2007). This study demonstrated the convergent validity of the Ohio Scales Problem Severity scale by using methods similar to previous research with more established measures such as the CBCL (Kasius et al. 1997; Rosenblatt and Rosenblatt 2002). The validity of these subscales can be furthered by comparing the subscales found on the Ohio Scales to the corresponding subscales on the CBCL.

Both the CBCL and Ohio Scales use parallel forms, which allow for both parents and youth to report on current

problems. A unique aspect of children’s mental health is that children are often brought to the attention of the mental health system by a parent or a teacher (Roth and Leavey 2006). Due to lack of insight into emotional and behavioral problems, younger children may not be reliable reporters, and as such, the parents report on a child’s symptoms. However, research has shown that child reports provide additional information that may be unknown to parents (e.g., anxiety, depression, lying; Kenny and Faust 1997). Due to the importance of multiple informants for child and adolescent mental health, the Ohio Scales created parallel forms so that both parents and youth could report on the severity of the child’s behavior. Even with identical forms, it is still necessary to examine whether youth and their parents report emotional and behavioral problems along similar dimensions. There is significant research that shows that youth and parents often do not agree about the child’s specific symptoms (*for review see* De Los Reyes and Kazdin 2005); however, it is important to determine whether youth and their parents have similar conceptual views of emotional and behavioral problems. In other words, regardless of the discrepancies in the reporting of symptom severity for an individual child, one would expect that youth and parents would group emotional and behavioral symptoms into similar over-arching constructs. This similar pattern in factor structure has been found in other measures of childhood emotional and behavioral problems such as with the CBCL/YSR (Achenbach and Rescorla 2004), SDQ (Muris et al. 2003) and the Behavior Assessment System for Children (BASC; Matazow and Kamphaus 2001). Thus, it would be expected that the factor structure for the Ohio Scales would also be consistent across informants.

The current study intended to contribute to the validity of the subscales of the Ohio Scales Problem Severity scale through exploring the factor structure and comparing these subscales to the corresponding subscales on the CBCL and YSR. Our study utilized exploratory factor analysis (EFA) to determine the initial factor structure of the Problem Severity scale, and then confirmed the factor structure through confirmatory factor analysis (CFA) utilizing a withheld sample. Although previous studies have shown a three-factor model as the best fit, based on an item-level comparison to a similar measure, it was hypothesized that a two-factor model may better fit the Ohio Scales Problem Severity scale for both the parent- and youth-report. Further, in order to support the concurrent validity of the hypothesized factors, these factors were compared to the well-established factors on the CBCL and YSR. Finally, the current study compared the factor structures of Ohio Scales Problem Severity scale for parent and youth report. It was hypothesized that the same factor structure would be consistent across both reporters.

Method

Participants

Participants included parent and youth dyads who received an intake evaluation between 1 January 2008 and 31 December 2012 at a CMH agency serving children, adolescents, and families in a semi-rural county in the Midwest. There were 1499 unique youth reports during this time period, of which 1269 (85 %) were matched with the corresponding parent reports to create parent and youth dyads. Of those 1269 dyads, 60 (4.72 %) were missing at least one item on the Ohio Scales Problem Severity scale and were not included in analyses. Additionally, five dyads (0.33 %) were removed as the youth were too young to complete the YSR. Finally, 119 dyads (7.94 %) were removed due to one parent reporting on multiple siblings, in order to prevent violations for independence, reducing the final sample to 1083 (72.25 %) unique youth and parent dyads. *T*-tests and χ^2 analyses were used to compare the included cases to the excluded cases across demographics and total scores on all measures; excluded youth scored significantly lower than included youth on the Internalizing ($t(1267) = -2.39$, $p = .017$) and Externalizing ($t(1267) = -2.17$, $p = .030$) subscales on the YSR. As a result, the current sample may represent more distressed youth than the full population of clients. The sample was approximately evenly split between males (50.5 %) and females (49.5 %), and the majority of the sample self-reported as Caucasian (88.1 %). The youth ranged in age from 11 to 18 years old, with a mean age of 14.74 years ($SD = 2.20$). The majority of the parent reports were from the biological mother of the child (78.7 %). The median household income in the county from which data was collected is US\$53,441 per year (U.S. Census Bureau 2014).

Procedure

We used archival data collected as part of the ongoing monitoring of a youth and family CMH center serving a semi-rural Midwest county. At this CMH, youth and families are either self-referred or referred by an outside agency (e.g., the public school system, law enforcement, hospital) that has identified the youth as needing mental health services. The agency serves youth for a variety of reasons, including acting out behavior, depression, anxiety, suicide attempts, and substance abuse. Prior to receiving services, the Ohio Scales and CBCL are both collected from all families during an intake assessment. Parents and guardians are asked to complete both the Ohio Scales and CBCL, and youth between the ages of 11 and 18 years old completed the parallel forms of these two measures. The Ohio Scales are also collected at 3 months, 6 months,

9 months, and 1 year after intake for all youth who continue to receive services at the CMH; for the current study, only the Ohio Scales completed at intake were used for the analyses. For youth who have been discharged and readmitted, and therefore have multiple intakes during the study time frame, only the first intake within the time frame was used. These data were collected and entered into database software as standard protocol for the diagnostic intake at the CMH.

Measures

Ohio Scales

The Ohio Scales-Short Form is a 48-item measure that includes scales for four domains: Problem Severity (20 items), Functioning (20 items), Hopefulness (4 items), and Satisfaction with Service (4 items). This was shortened from the original form, which contained 44 items on the Ohio Scales Problem Severity scale, 20 items on the Functioning scale, 4 items on the Hopefulness scale, and 4 items on the Satisfaction with Services scale (Ogles et al. 2000). Based on feedback from service providers and a factor analysis of the original 48 items on the Ohio Scales Problem Severity scale, the Ohio Scales Problem Severity scale was shortened to 20 items and other items were reworded for consistency across the three forms, producing the current 48-item short form (Ogles et al. 2000). The short form is highly correlated with the original longer form ($r = .80-.96$; Ogles et al. 2000). The Ohio Scales Problem Severity scale contains items addressing many different emotional and behavior problems (e.g. “Arguing with others”, “Feeling worthless or useless”), and the frequency of these problems occurring in the past month are rated on a 6-point scale from 0 (*Not at All*) to 5 (*All of the Time*). The Ohio Scales Problem Severity scale is scored by summing the 20 items for a total severity score ranging from 0 to 100, with a clinical cutoff of 25. No items are reverse-scored.

The Ohio Scales use three parallel forms to gather information from different sources, including youth (ages 9–18), parents, and workers (clinical service providers). However, because the goal of the current study was to compare the Ohio Scales to the CBCL, only the parent and youth reports on the Ohio Scales Problem Severity scale were used. The CBCL has a parallel teacher report form, but it does not have a comparable worker form; thus, no comparison can be made between the worker form of the Ohio Scales and CBCL.

The Ohio Scales Problem Severity scale has been shown to be valid and reliable both in the longer (44-item) and shorter (20-item) versions. The shorter version has also demonstrated excellent internal consistency for both parent and youth reports ($\alpha = .91$ and $.92$, respectively; TDMHMR 2004).

Similarly, internal consistency in the current sample was also high for both parents and youth ($\alpha = .88$ and $.88$, respectively). Additionally, both the youth-report and parent-report Ohio Scales Problem Severity scales have been significantly correlated with well-established measures of childhood mental health and behavior problems such as the YSR and CBCL ($r = .62$ and $.64$, respectively) and the youth-report and parent-report SDQ ($r = .56$ and $.63$, respectively; TDMHMR 2004), supporting convergent validity. The Ohio Scales Problem Severity scale has also demonstrated construct validity, with clinical samples receiving significantly higher scores on the scale than community samples (Dowell and Ogles 2008).

The Achenbach CBCL

The Achenbach CBCL (Achenbach 1991) is a well-established and widely-used measure of child and adolescent mental health and problem behaviors. The CBCL is a 113-item parent report of child behavior which assesses various emotional and behavioral problems. The 113 items are rated on a 3-point scale from 0 (*Not True*) to 2 (*Very True or Often True*) based on the past 6 months. There is also a parallel self-report form for youth called the Youth Self Report (YSR). The 113 items are summed and then converted to *T*-scores based on age and gender to create a Total Problems score. Additionally, the CBCL contains several subscales, including eight Syndrome Scales, six DSM-oriented scales, and two broad Internalizing (e.g., anxiety, depression) and Externalizing (e.g., aggressive behavior, rule breaking) subscales. Each of these scales has demonstrated adequate psychometric properties (Achenbach 1991; Achenbach and Rescorla 2004).

Data Analyses

Exploratory Factor Analysis

Before beginning the primary analyses, the dataset was randomly split into two relatively equivalent subsamples. An EFA was performed on one subsample, and then the resulting factor structure was tested using a CFA on the other subsample. Due to possible differences in factor loadings for individual items, separate EFAs were used for parent and youth reports. The factor structure of the 20 items of the Ohio Scales Problem Severity scale was examined with an EFA using principal axis factoring extraction with promax (oblique) rotation applied to the solution. An oblique rotation was used because it allows for factors within a model to be correlated and previous research has shown that Externalizing and Internalizing factors correlate (Achenbach and Rescorla 2004). To select the optimal number of factors, we examined the scree plot,

eigenvalues over one (Kaiser 1960), and the factor loadings over .50 (Hair et al. 1998) in the Rotated Pattern matrix. All items were retained, even those with low factor loadings and/or cross-loadings on multiple factors. The Ohio Scales Problem Severity scale is currently being used in many state-wide public health systems, so to facilitate generalizability of findings we decided not to alter the structure of the scale.

Confirmatory Factor Analysis

CFA was used to test the factor structures established using the EFA. As with the EFA, separate CFAs were used for the youth and parent reports due to possible differences in strengths of factor loadings for individual items. Model fit was determined by examining several estimations of fit, including: χ^2 indices, Root Mean Square Error of Approximation (RMSEA), Root-Mean-Square Residual (RMR), and the Comparative Fit Index (CFI). While non-significant χ^2 remains the gold standard for model fit, with very large samples such as in the our study, χ^2 's are likely to be inflated and therefore incremental fit indices can be more useful (Cheung and Rensvold 2002). Rules of thumb for adequate model fit using incremental fit indices include a cutoff of below .06 for RMSEA, below to .08 for RMR, and above .95 for CFI (Hu and Bentler 1999). Additionally, a multiple-groups analysis was run to establish factor structure invariance across the youth and parent samples.

Concurrent Validity with CBCL

Once a factor structure was confirmed, structural equation modeling was used to test if the subscales derived from the Ohio Scales Problem Severity parent-reports and youth-reports correspond to the related subscales on the CBCL and YSR. Separate models were completed for the youth-report and parent-report Ohio Scales and the YSR and CBCL, respectively.

Data Cleaning and Missing Value Analysis

In order to assess reliability of data entry, case records were pulled for approximately 10 % ($n = 122$) of the total sample to compare original paper forms to the responses entered into the database. This reliability check uncovered a consistent pattern of missing items being entered as 0 on all of the measures. In total, 116 items out of a possible 32,330 (<1 %) were entered as 0 in the database but were left blank by the respondent. These erroneous 0's were deleted from the data, creating a "clean" subsample in which items left blank were missing in the dataset. Using the clean subsample, Little's test for missing completely at random was performed separately for each measure (Little 1988).

Results of Little's test indicated that the missing data were missing completely at random.

Expectation maximization (Schlomer et al. 2010) was used to impute the missing data for cases that had less than eight items missing on the CBCL and YSR (7 % of total items) and less than 5 items missing on Ohio Scales Youth and Ohio Scales Parent (25 % of total items). Of the 122 cases in the subsample, four cases (3 %) were dropped due to having more than eight items missing on the CBCL or YSR. Of the remaining 118 cases, 19 % had at least one item imputed on the Ohio Scales Youth or Ohio Scales Parent, and 42 % had at least one item imputed on the CBCL or YSR. After imputing the missing values, correlations between measures were compared between the cleaned subsample with imputed values and original subsample with the erroneous zeros. No significant differences were found in the correlations between measures in the two datasets, suggesting that there were no biases due to the erroneous zeros. Together, these missing data analyses suggest that respondents may have often left items blank when a symptom was not present, resulting in minimal differences in the data when those blanks were erroneously entered as zeros.

Given the minimal impact of these erroneous missing values, the full sample was used. Miss data analysis on the

full sample indicated that missing values on the Ohio Scales Problem Severity scale, CBCL, and YSR were missing completely at random and thus listwise deletion was utilized to remove the 60 cases with missing data on these measures.

Results

Prior to running any analyses, the sample was randomly split into two subsamples, an EFA subsample and CFA subsample. Means and standard deviations for all items for both parent and youth reports can be found in Table 1. Demographics and total scores were compared between the two samples to ensure that random split provided equivalent samples (Table 2). The two samples did not significantly differ on gender, age, ethnicity, parent reporting, or mean scores on either the Ohio Scales or CBCL/YSR.

Exploratory Factor Analysis

Parent Report

In the initial EFA, using principal axis factoring extraction with promax (oblique) rotation, the Kaiser's rule indicated that four factors be extracted, explaining 45.92 % of the

Table 1 Means and standard deviations for all items

Item	Parent		Youth	
	EFA (<i>n</i> = 540)	CFA (<i>n</i> = 543)	EFA (<i>n</i> = 540)	CFA (<i>n</i> = 543)
Arguing with others	2.37 (1.38)	2.29 (1.41)	2.13 (1.30)	2.08 (1.32)
Getting into fights	0.72 (1.08)	0.77 (1.21)	0.77 (1.11)	0.70 (1.05)
Yelling, swearing, or screaming at others	1.76 (1.44)	1.87 (1.53)	1.59 (1.46)	1.55 (1.43)
Fits of anger	1.69 (1.44)	1.65 (1.50)	1.40 (1.43)	1.46 (1.48)
Causing trouble for no reason	0.89 (1.24)	0.94 (1.35)	0.66 (1.07)	0.72 (1.12)
Refusing to do things teachers or parents ask	1.84 (1.55)	1.85 (1.58)	1.21 (1.28)	1.24 (1.35)
Using drugs or alcohol	0.29 (0.77)	0.31 (0.79)	0.39 (0.97)	0.36 (0.93)
Breaking rules or breaking the law (out past curfew, stealing)	0.79 (1.22)	0.71 (1.18)	0.54 (1.00)	0.55 (1.03)
Skipping school or classes	0.21 (0.72)	0.25 (0.79)	0.21 (0.69)	0.26 (0.81)
Lying	1.57 (1.50)	1.46 (1.50)	1.36 (1.27)	1.37 (1.31)
Can't seem to sit still, having too much energy	1.14 (1.47)	1.10 (1.53)	1.81 (1.72)	1.57 (1.67)
Talking or thinking about death	0.54 (0.94)	0.59 (1.01)	0.74 (1.28)	0.76 (1.23)
Hurting self (cutting or scratching self, taking pills)	0.35 (0.83)	0.35 (0.81)	0.43 (1.03)	0.43 (0.95)
Feeling worthless or useless	1.06 (1.28)	0.99 (1.29)	1.25 (1.53)	1.23 (1.51)
Feeling lonely and having no friends	1.12 (1.38)	1.14 (1.38)	1.08 (1.45)	1.03 (1.47)
Feeling anxious or fearful	1.01 (1.28)	1.02 (1.35)	1.07 (1.38)	1.10 (1.43)
Worrying that something bad is going to happen	0.86 (1.24)	0.91 (1.30)	1.52 (1.59)	1.50 (1.59)
Feeling sad or depressed	1.65 (1.40)	1.66 (1.44)	1.71 (1.62)	1.67 (1.67)
Nightmares	0.48 (0.96)	0.54 (1.06)	1.04 (1.52)	0.98 (1.42)
Eating problems	0.64 (1.25)	0.69 (1.31)	0.81 (1.43)	0.84 (1.44)

Table 2 Demographics and total scores for youth

	Full (N = 1083)	EFA (n = 540)	CFA (n = 543)
Gender			
Male	536 (49.5 %)	265 (49.1 %)	271 (49.9 %)
Female	547 (50.5 %)	275 (50.9 %)	272 (50.1 %)
Race/Ethnicity			
White	954 (88.1 %)	470 (87.0 %)	484 (89.1 %)
Hispanic	32 (3.0 %)	16 (3.0 %)	16 (2.9 %)
African-American	19 (1.7 %)	14 (2.6 %)	5 (0.9 %)
Other	78 (7.2 %)	40 (7.4 %)	38 (7.0 %)
Mean age (SD)	14.74 (2.20)	14.61 (2.27)	14.86 (2.12)
11–13 years	370 (34.2 %)	205 (38.0 %)	165 (30.4 %)
14–16 years	552 (51.0 %)	258 (47.8 %)	294 (54.1 %)
17–18 years	161 (14.9 %)	77 (14.3 %)	84 (15.5 %)
Parent Ohio Scales problem severity (SD)	21.04 (13.98)	20.99 (13.28)	21.08 (14.65)
Youth Ohio Scales problem severity (SD)	21.56 (14.97)	21.72 (14.41)	21.41 (15.52)
CBCL internalizing (SD)	13.07 (9.00)	14.20 (8.74)	13.93 (9.26)
CBCL externalizing (SD)	17.36 (11.77)	17.53 (11.26)	17.20 (12.27)
YSR internalizing (SD)	17.62 (11.24)	17.88 (11.38)	17.37 (11.11)
YSR externalizing (SD)	16.01 (9.36)	16.17 (9.15)	15.85 (9.57)

* $p < .05$

variance. The pattern matrix showed that three items (i.e., Breaking Rules, Energy, and Eating) did not have a factor loading at or above .50 on any of the four factors, but only one of those items had a factor loading below .40 (i.e., Eating). Additionally, one item (i.e., Breaking Rules) cross-loaded above .30 on two factors. Examining the items on each factor (Table 3), it appears that the first factor contains items related to *Aggressive Behavior* (27.35 % variance explained), the second factor represents *Depression* (11.99 % variance explained), the third factor represents *Anxiety* (3.48 % variance explained), and the fourth factor represents *Delinquent Behavior* (3.10 % variance explained). Given that the fourth factor only explained 3.10 % of the variance, a second EFA was run in which the factor extraction was limited to three factors, essentially replicating previous factor analyses of the Ohio Scales Problem Severity scale (TDMHMR 2004).

This second EFA, which was limited to three-factors, accounted for 42.55 % of the variance, slightly less than the four-factor model. The pattern matrix indicated that five items had factor loadings below .50 (i.e., Drugs, Skipping, Energy, Nightmares, and Eating), and three of those items had factor loadings below .40 (i.e., Drugs, Skipping, and Eating). No items cross-loaded above .30 on two factors. Examining the items on each factor (Table 3), it appears that the first factor represents *Externalizing* (27.24 % variance explained), the second factor represents *Depression* (11.88 % variance explained), and the third factor represents *Anxiety* (3.44 % variance explained). Further determinations

of the appropriate factor structure were explored by testing both the three-factor and four-factor models using CFA.

Finally, to test the hypothesized two-factor model, an EFA was run in which the factor extraction was limited to two factors. The two-factor model accounted for 38.70 % of the variance, which was less than both the three-factor and four-factor model. The pattern matrix indicated that six items had factor loadings below .50 (i.e., Drugs, Skipping, Energy, Nightmares, Eating, and Hurting), and four of those items had factor loadings below .30 (i.e., Skipping, Energy, Eating, and Hurting) on all factors. It should be noted that no items had a cross-loading of .30 or greater on more than one factor. Examining the items on each factor, it appears that the first factor represents *Externalizing* (27.10 % variance explained) and the second factor represents *Internalizing* (11.60 % variance explained). The two-factor EFA appeared to be a worse fit than both the three-factor and four-factor models, and thus it was not further explored using CFA.

Youth Report

Using principal axis factoring extraction with promax (oblique) rotation and the Kaiser’s rule to determine the number of factors to be extracted, the unrestrained EFA indicated that five factors were to be extracted, explaining 48.12 % of the variance. The pattern matrix revealed that five items (i.e., Eating, Worthless, Energy, Lying, and Skipping) did not load at or above .50 on any of the four

Table 3 Factor loadings for parent report EFA

Item	Three-factor model			Four-factor model			
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
Arguing with others	.75	-.03	.16	.88	-.05	.08	-.13
Getting into fights	.55	.02	.03	.56	.01	.00	.01
Yelling, swearing, or screaming at others	.75	-.01	.12	.80	-.02	.06	-.04
Fits of anger	.66	.18	.07	.74	.18	.00	-.09
Causing trouble for no reason	.78	-.06	.01	.83	-.06	-.04	-.04
Refusing to do things teachers or parents ask	.78	.02	-.06	.77	.02	-.09	.02
Using drugs or alcohol	.31	.01	-.16	-.06	.05	.00	.57
Breaking rules or breaking the law (out past curfew, stealing)	.67	.02	-.16	.38	.04	-.04	.47
Skipping school or classes	.29	-.04	-.04	-.04	-.03	.13	.54
Lying	.66	.05	-.08	.52	.05	-.04	.21
Can't seem to sit still, having too much energy	.31	-.21	.40	.31	-.22	.40	.07
Talking or thinking about death	.06	.60	-.32	-.02	.58	-.28	.07
Hurting self (cutting or scratching self, taking pills)	-.01	.58	.05	-.02	.60	.06	.02
Feeling worthless or useless	.00	.77	.07	.03	.77	.06	-.03
Feeling lonely and having no friends	-.02	.60	.11	.04	.59	.08	-.09
Feeling anxious or fearful	-.10	.09	.78	-.04	.08	.79	.03
Worrying that something bad is going to happen	-.16	.02	.80	-.06	.01	.77	-.03
Feeling sad or depressed	-.03	.66	.19	-.02	.66	.20	.01
Nightmares	.02	-.06	.48	.04	-.06	.50	.05
Eating problems	.12	.18	.14	.07	.18	.17	.09

Note: Bold indicates strongest factor loading for each item

factors, and three of those four items (i.e., Energy, Lying, and Eating) loaded below .40 on all factors. One item (i.e., Worthless) had a cross-loading of .30 or greater on more than one factor. Additionally, the fifth factor had only one item with a factor loading greater than .50. Given the poor factor loadings on the fifth factor, it was determined that this factor was not valid, and thus the five-factor model was not further interpreted.

A second EFA was run in which the factor extraction was limited to four factors. The EFA indicated that four factors explained 45.82 % of the variance. The pattern matrix revealed that five items (i.e., Energy, Worthless, Lying, Skipping, and Eating) did not load at or above .50 on any of the four factors, but two of those five items (i.e., Worthless and Lying) loaded above .40 on at least one factor. One item (i.e., Worthless) had a cross-loading of .30 or greater on more than one factor. Examining the items on each factor (Table 4), it appears that the first factor contains items related to *Internalizing* (27.99 % variance explained), the second factor represents *Externalizing* (11.03 % variance explained), the third factor represents *Suicidal* (3.87 % variance explained), and the fourth factor represents *Delinquent Behavior* (2.94 % variance explained). Given that one item cross-loaded on two factors and that the fourth

factor only explained 2.94 % of the variance, a three-factor model may be a better fit for the data.

The three-factor model accounted for 42.39 % of the variance, slightly less than the four-factor model. The pattern matrix indicated that five items had factor loadings below .50 (i.e., Hurting, Eating, Lying, Energy, and Skipping), but only one of those five items (i.e., Energy) loaded below .40 on all factors. It should be noted that no items had a cross-loading of .30 or greater on more than one factor. Examining the items on each factor (Table 4), it appears that the first factor represents *Internalizing* (27.78 % variance explained), the second factor represents *Externalizing* (10.87 % variance explained), and the third factor represents *Delinquent Behaviors* (3.73 % variance explained). Further determinations of the appropriate factor structure were explored by testing both the three-factor and four-factor models using CFA.

Finally, to test the hypothesized two-factor model, an EFA was run in which the factor extraction was limited to two factors. The two-factor accounted for 38.40 % of the variance, which was less than both the three-factor and four-factor model. The pattern matrix indicated that five items had factor loadings below .50 (i.e., Drugs, Skipping, Energy, Hurting, and Eating), and three of those items had

Table 4 Factor loadings for youth report EFA

Item	Three-factor model			Four-factor model			
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
Arguing with others	.08	.72	-.13	.08	.72	-.01	-.13
Getting into fights	.02	.56	-.03	-.04	.57	.07	-.05
Yelling, swearing, or screaming at others	-.03	.76	-.01	-.01	.76	-.03	-.01
Fits of anger	.07	.70	-.08	.22	.68	-.20	-.04
Causing trouble for no reason	-.06	.62	.12	-.13	.64	.08	.09
Refusing to do things teachers or parents ask	-.06	.74	.02	-.08	.74	.02	.00
Using drugs or alcohol	.02	-.10	.65	.01	-.11	.06	.64
Breaking rules or breaking the law (out past curfew, stealing)	-.03	.17	.62	.05	.13	-.06	.69
Skipping school or classes	.04	-.02	.40	-.01	-.01	.09	.38
Lying	.01	.44	.24	-.08	.45	.12	.22
Can't seem to sit still, having too much energy	.11	.29	-.01	.14	.28	-.04	.00
Talking or thinking about death	.67	-.01	.11	.19	.03	.67	.01
Hurting self (cutting or scratching self, taking pills)	.50	-.06	.16	.03	-.03	.63	.07
Feeling worthless or useless	.81	-.01	.00	.49	.02	.44	-.05
Feeling lonely and having no friends	.74	-.05	.03	.54	-.03	.27	.01
Feeling anxious or fearful	.73	.03	-.04	.88	-.02	-.12	.05
Worrying that something bad is going to happen	.73	.01	-.05	.90	-.04	-.16	.05
Feeling sad or depressed	.80	.03	-.02	.65	.04	.21	-.02
Nightmares	.51	.03	-.06	.54	.01	-.02	-.02
Eating problems	.45	.09	-.03	.33	.10	.15	-.04

Note: Bold indicates strongest factor loading for each item

factor loadings below .40 (i.e., Drugs, Skipping, and Energy) on all factors. Additionally, six items (i.e., Sad, Worthless, Anxious, Death, Anger, and Arguing) cross-loaded above .30 on two factors. Examining the items on each factor, it appears that the first factor represents *Internalizing* (27.68 % variance explained) and the second factor represents *Externalizing* (10.71 % variance explained). The two-factor EFA appeared to be a worse fit than both the three-factor and four-factor models, and thus it was not further explored using CFA.

Confirmatory Factor Analysis

Parent Report

Due to the inconclusive results of the EFA, CFA was used to compare both the three-factor and the four-factor models using the withheld subsample of parent reports. First, the three-factor model identified in the parent EFA was tested using the withheld sample of parent reports. The three-factor CFA specified three latent factors: *Externalizing*, *Anxiety*, and *Depression*. Multiple fit indices indicated that the three-factor model provides a moderately acceptable fit to the data (Table 5). The χ^2 was significant (χ^2 (167) = 876.60, $p < .001$); however, RMSEA suggests borderline-

acceptable fit (RMSEA = .091; CI 90 % = .085–.096). Other incremental fit indices (SRMR = .09, CFI = .93) also suggested a borderline-acceptable fit for the three-factor model.

Next, the four-factor model identified in the parent EFA was tested using the withheld sample of parent reports. The four-factor CFA specified four latent factors: *Aggression*, *Delinquency*, *Depression*, and *Anxiety*. Multiple fit indices indicated that the four-factor model appears to be an acceptable fit to the data. Although, the χ^2 was significant (χ^2 (164) = 647.00, $p < .001$), other incremental fit indices indicated an adequate fit to the data. RMSEA suggested an acceptable fit (RMSEA = 0.076; CI 90 % = .070 –.082). Additionally, other incremental fit indices indicated a good fit (SRMR = .06, CFI = .96). Model fit between the three-factor and four-factor models (Table 5) were compared using a χ^2 difference test, which indicated that the four-factor model was a significantly better fit than the three-factor model ($\Delta\chi^2$ (3) = 229.60, $p < .001$). Thus, the four-factor model was retained as the best fitting model. The final four-factor model with parameter estimates can be seen in Fig. 1.

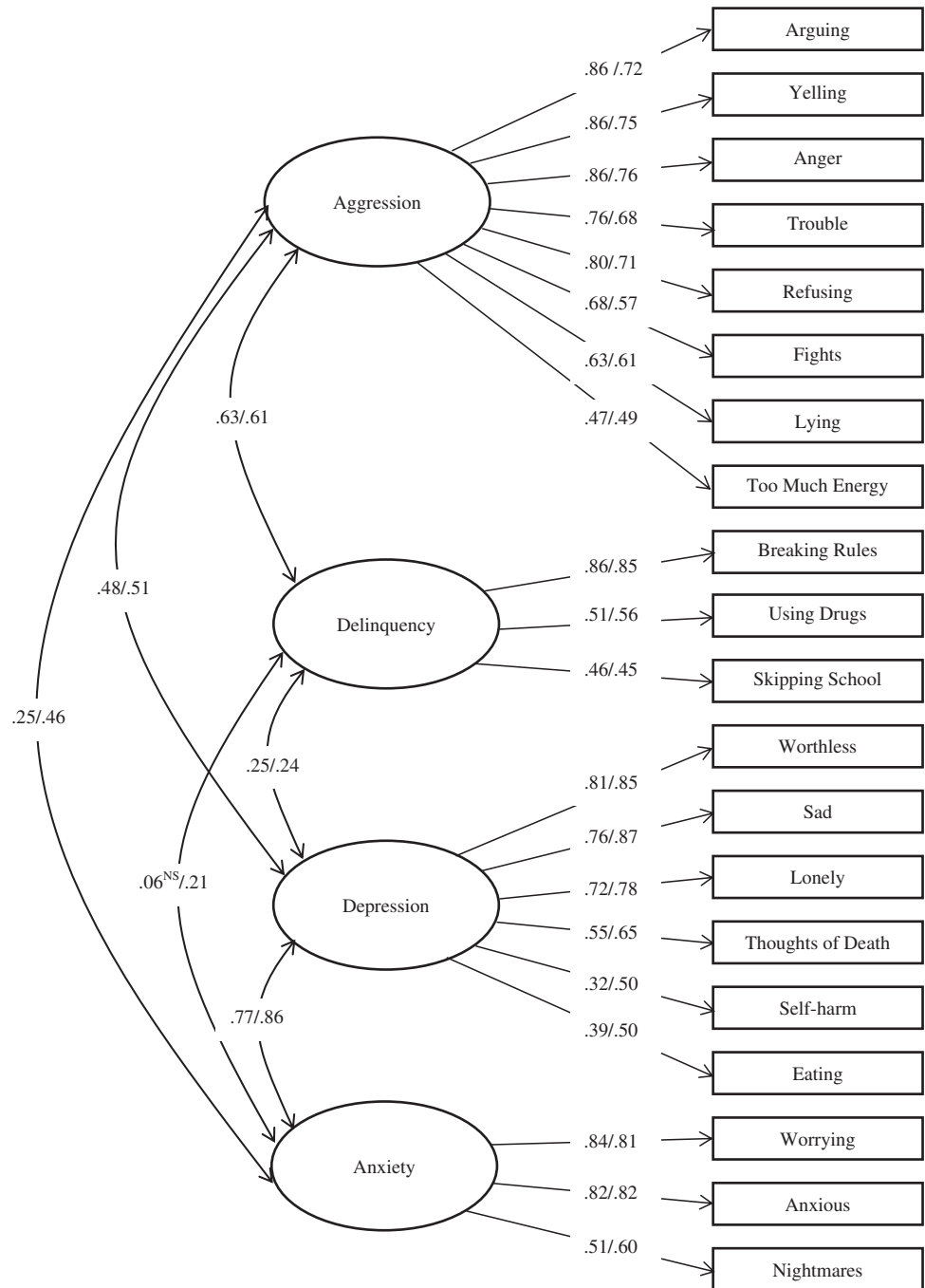
Factor loadings for all four latent variables in the CFA were significant. Variance explained suggested that six indicators had less than 30 % of their variance explained by

Table 5 Comparison between three-factor and four-factor models for parent report

	χ^2 (df)	RMSEA (95 % CI)	CFI	SRMR	$\Delta\chi^2$ (df)
3 Factor model	876.60 (167)*	0.091 (0.085–0.096)	0.93	0.09	
4 Factor Model	647.00 (164)*	0.076 (0.070–0.082)	0.96	0.06	229.60 (3)*

* $p < .001$

Fig. 1 Four-factor structure of Ohio Scales for parent/child report. Note: All estimates are standardized. All pathways are significant unless noted with ns



the corresponding latent variable (Drugs 26 %; Skip 21 %; Energy 23 %; Hurt 10 %; Nightmare 26 %; Eating 16 %); the remaining fourteen variables had between 30–74 % of

their variance explained. Several of the latent variables were significantly correlated—Aggression and Anxiety ($r = .25$; $p < .05$), Aggression and Depression ($r = .48$; $p < .05$),

Aggression and Delinquency ($r = .63; p < .05$), Anxiety and Depression ($r = .77; p < .05$), and Depression and Delinquency ($r = .25; p < .05$). Only Anxiety and Delinquency were not significantly correlated ($r = .06$; ns).

Youth Report

As with the parent report, CFA was used to compare both the three-factor and four-factor models using the withheld subsample of youth reports. The three-factor model identified in the youth EFA was tested using the withheld sample of youth reports. The three-factor CFA specified three latent factors: *Externalizing*, *Internalizing*, and *Delinquent Behaviors*. Multiple fit indices indicated that the three-factor model appears to be an adequate fit to the data (Table 6). The χ^2 was significant ($\chi^2 (167) = 685.31, p < .001$), but other incremental fit indices indicated a more acceptable model fit. RMSEA suggested borderline-acceptable fit (RMSEA = .080; CI 90 % = .074–.086), but other incremental fit indices indicated a good fit (SRMR = .06; CFI = .95).

The four-factor model identified in the youth EFA was tested using the withheld sample of youth reports. The four-factor CFA specified four latent factors: *Internalizing*, *Externalizing*, *Suicidal*, and *Delinquent Behaviors*. Multiple fit indices indicated that the four-factor model appears to be an adequate fit to the data (Table 6). The χ^2 was significant ($\chi^2 (164) = 625.59, p < .001$), but RMSEA and other fit indices suggested an acceptable fit (RMSEA = 0.076, CI 90 % = .070–.082; SRMR = .06; CFI = .96). Model fit between the three-factor model and youth EFA-derived four-factor model was compared using a χ^2 difference test which indicated that the four-factor model was a significantly better fit than the three-factor model ($\Delta\chi^2 (3) = 59.72, p < .001$).

Complicating analyses of the models of the youth reported data, the four-factor model identified by the EFA for youth reports differed from the four-factor model for parent reports. Specifically, the anxious and depressed items from the parent EFA loaded together in the youth EFA, except for the self-harm and thoughts of death items, which formed a single factor for the youth EFA. In order to test if the four-factor model identified for parent report would also fit for youth report, a final CFA model was run to replicate the parent report model with the youth reports. This final model replicated the structure of the parent four-factor model using the withheld sample of youth reports. The CFA specified four latent factors: *Aggression*, *Delinquency*, *Depression*, and *Anxiety*. Multiple fit indices indicated that the four-factor model appears to be a good fit to the data. Although, the χ^2 was significant ($\chi^2 (164) = 611.82, p < .001$), other incremental fit indices indicated a good fit to the data. RMSEA suggests acceptable fit (RMSEA

= .074; CI 90 % = .068–.080). Additionally, other fit indices indicated a good fit (SRMR = .06, CFI = .96). Model fit between the three-factor and four-factor parent replication model was compared using a χ^2 difference test which indicated that the four-factor parent replication model was a significantly better fit than the ($\Delta\chi^2 (3) = 73.49, p < .001$). The two four-factor models could not be compared using a χ^2 difference test as they have the same degrees of freedom. However, given the lower χ^2 value and slightly lower RMSEA, it can be reasoned that the parent-replicated four-factor model for the youth report is as good or a slightly better fit than the original four-factor model. Additionally, having a consistent factor structure across both parent and youth reports contributes to a more parsimonious measure for the Ohio Scales. Thus, the parent-replicated four-factor model was retained as the best-fitting model for the youth-report Ohio Scales. This final four-factor model with parameter estimates can be seen in Fig. 1.

Factor loadings for all four latent variables in the CFA were significant. Variance explained suggested that four indicators had less than 30 % of their variance explained by the corresponding latent variable (Energy 24 %; Hurt 25 %; Eating 25 %; and Skipping 20 %); the remaining sixteen variables had between 32–76 % of variance explained. All of the latent variables were significantly correlated; Aggression and Anxious ($r = .46; p < .05$), Aggression and Depression ($r = .51; p < .05$), Aggression and Delinquency ($r = .61; p < .05$), Anxious and Depression ($r = .86; p < .05$), Anxious and Delinquency ($r = .21; p < .05$), and Depression and Delinquency ($r = .24; p < .05$).

Multiple Groups Analysis

After establishing a four-factor model for parents and then replicating a comparable model for youth, multiple-group models were tested to look for invariance in the measurement model of the Ohio Scales Problem Severity scale between parent and youth reports. Three increasingly stringent levels of invariance were tested: configural, metric, and scalar. To determine if models met criteria at each level of invariance, the global fit statistics— χ^2 , RMSEA, and CFI—for the multiple-group model were examined. It should be noted that when testing multiple-group models, it is common to find a significant χ^2 , and thus other fit indices may be better indicators of overall model fit (Cheung and Rensvold 2002). Global fit indices for the test of configural invariance—meaning that observed variables load onto the same latent variables for both groups—indicated that the overall structure of the two measurement models were equivalent ($\chi^2 (328) = 1301.85, p < .001$; RMSEA = .077; CI 90 % = .073–.081; CFI = .96). Global fit indices for the test of metric invariance—tested by constraining the factor loadings of the same items to be equal across both

Table 6 Comparison between three-factor and four-factor models for youth report

	χ^2 (df)	RMSEA (95 % CI)	CFI	SRMR	$\Delta\chi^2$ (df) with 3 factor model
3 Factor model	685.31 (167)*	0.080 (0.074–0.086)	0.95	0.06	
EFA 4 factor model	625.59 (164)*	0.076 (0.070–0.082)	0.96	0.06	59.72 (3)*
Parent 4 factor model	611.82 (164)*	0.074 (0.068–0.080)	0.96	0.06	73.49 (3)*

* $p < .001$

models—suggested that the model provided adequate fit to the data (χ^2 (344) = 1581.74, $p < .001$; RMSEA = .084; CI 90 % = .080–.088; CFI = .94). The fit statistics of the metric invariance model were compared to those of the configural invariance model, and the difference of .02 for the CFI and .007 for RMSEA suggested that fit was not drastically different between the two models (Cheung and Rensvold 2002). Thus, it can be assumed that the model has both configural invariance and metric invariance. Global fit indices for the test of scalar invariance—tested by constraining the intercepts for the observed variables to be equal across the two models—suggested that the model does not provide an adequate fit to the data (χ^2 (364) = 2144.24, $p < .001$; RMSEA = .097; CI 90 % = .093–.10; CFI = .92). Thus, it appears that the four-factor model of the Ohio Scales Problem Severity scale did not meet the criteria for scalar invariance. Taken together, these three test of invariance suggest that the factors have the same items across parents and youth (i.e., configural invariance) and that the factor loadings for this items of similar across parent and youth models (i.e., metric invariance); however, the intercepts and factor loadings together are not similar across models (i.e., scalar invariance). From a practical standpoint this suggest that although the constructs in the four-factor model are similar for parent and youth report, the scores on these latent factors are not directly comparable (for more detailed discussion of types of measurement invariance, see Brown 2015).

Concurrent Validity of Ohio Scales Problem Severity with CBCL

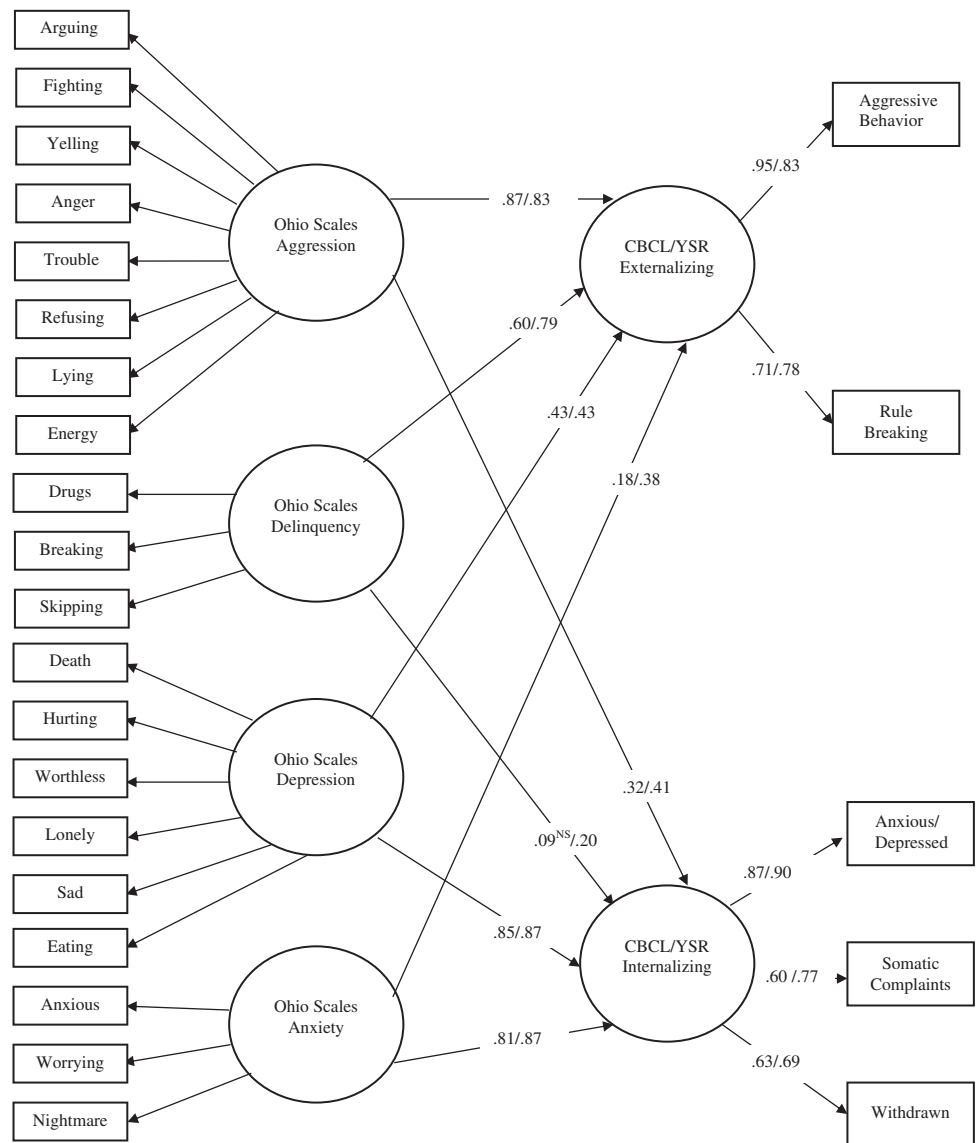
In order to test the concurrent validity of the Ohio Scales Problem Severity scale to the CBCL, structural equation modeling was used to compare similar factors from the two measures. The four factors found in the Ohio Scales Problem Severity scale are similar to the first-level latent variables that load onto the broad higher-order subscales of Internalizing and Externalizing on the CBCL and YSR. The Internalizing subscale of the CBCL and YSR is composed of three first-order subscales: Withdrawn-Depressed, Somatic Complaints, and Anxious-Depressed (Fig. 2). Two of the similar subscales found in the factor analyses of the Ohio Scales Problem Severity scale, Depression and

Anxiety, were hypothesized to be strongly correlated with the Internalizing subscale on the CBCL and YSR. The Externalizing subscale for the CBCL and YSR is composed of two first-order subscales, Aggressive Behavior and Delinquent Behavior. Two of the corresponding subscales were found in the factor analyses of the Ohio Scales Problem Severity scale and were hypothesized to be strongly correlated with the Externalizing subscale on the CBCL and YSR. Separate SEM models were run comparing the parent-report Ohio Scales Problem Severity scale to the CBCL and the youth-report Ohio Scales Problem Severity to the YSR.

The SEM model for the parent-report Ohio Scales Problem Severity and CBCL showed strong correlations between the theoretically similar subscales and weaker correlations for non-similar subscales, supporting the construct validity of the Problem Severity scale (Fig. 2). Specifically, the parent-report Ohio Scales Problem Severity Delinquency and Aggression subscales were strongly correlated with the CBCL Externalizing subscale ($r = .60$, $p < .05$ and $r = .87$, $p < .05$, respectively), but showed weaker correlations with the CBCL Internalizing subscale ($r = .09$, ns and $r = .32$, $p < .05$, respectively). Conversely, the parent report Ohio Scales Problem Severity Depression and Anxiety were strongly correlated with the CBCL Internalizing ($r = .85$, $p < .05$ and $r = .81$, $p < .05$, respectively), but showed weaker correlations with the CBCL Externalizing subscale ($r = .43$, $p < .05$ and $r = .18$, $p < .05$, respectively).

The youth-report Ohio Scales Problem Severity scale and the YSR demonstrated a similar pattern of correlations among the Ohio Scales Problem Severity subscales and the higher-order YSR subscales. Specifically, the youth-report Ohio Scales Problem Severity Aggression and Delinquency subscales were strongly correlated with the YSR Externalizing subscale ($r = .83$ and $r = .79$, respectively), but showed weaker correlations with the YSR Internalizing subscale ($r = .41$ and $r = .20$, respectively). Conversely, the youth-report Ohio Scales Problem Severity Depression and Anxiety subscales were strongly correlated with the YSR Internalizing ($r = .87$ and $r = .87$, respectively), but showed weaker correlations with the YSR Externalizing subscale ($r = .43$ and $r = .38$, respectively).

Fig. 2 Model testing associations between Ohio Scales Problem Severity and CBCL for Parent/Youth. Note: All estimates are standardized. All pathways are significant unless noted with NS



Discussion

The recent increase in demands for evidence of treatment outcomes from Community Mental Health organizations (CMHs) has led to the need for a practical measure of mental health outcomes (Dowell and Ogles 2008; Trask and Garland 2011). The Ohio Scales are a set of measures designed specifically to meet the needs for outcome measurement in CMHs (Ogles et al. 2001). The Ohio Scales provide a short and affordable measure with parallel forms for youth, parents, and service providers to report on four domains: Problem Severity, Hopefulness, Satisfaction with Service, and Functioning. Of primary interest in our study was youth and parent report on the Ohio Scales Problem Severity scale (Ohio Scales Problem Severity), a measure of child emotional and behavioral problems. The Ohio Scales

Problem Severity subscale has demonstrated acceptable psychometric properties and has been correlated with other well-established measures of child mental health (Ogles et al. 2001; TDMHMR 2004; Warnick et al. 2009). However, unlike other measures of child emotional and behavioral problems, the Ohio Scales Problem Severity scale lacks well-established subscales that are often found in similar measures. Our study established the factors within of the Ohio Scales Problem Severity and compared them to similar factors on the CBCL and Youth Self Report (YSR), the gold standards of parent-reports and youth-reports of child emotional and behavioral problems.

In a previous non-peer reviewed manuscript, the Ohio Scales Problem Severity was found to have three factors: Internalizing, Externalizing, and Delinquency (TDMHMR 2004). Due to the limitations of that study and the factor

structure of other measures of child emotional and behavioral problems (e.g., CBCL), we hypothesized that a two-factor model, Internalizing and Externalizing, would provide a better fit for the Ohio Scales Problem Severity scale. However, results from an EFA did not support the hypothesized two-factor model, but rather found that a four-factor model was appropriate. Although a four-factor model was found to be the best fit for both the youth and parent reports in the EFA, the items loading on to each of the four factors differed between the two groups.

Unlike previous factor analyses of the Ohio Scales Problem Severity (Ogles et al. 2000; TDMHMR 2004), the current study attempted to validate the four-factor models found in the EFA, using a randomly-split, withheld sample to conduct a CFA. Because EFA is an exploratory approach, it is less guided by theory, and thus factor structures found in EFAs are primarily based on statistical covariance. However, CFA provides the flexibility to test and compare various factor structures, thus allowing empirically-derived factor structures from the EFA to be compared to more theoretically-based factor structures. The CFA confirmed the four-factor model found in the EFA for parent reports. However, for youth report, the four-factor model found in the youth EFA was not the best fitting model when compared to an alternate four-factor model which replicated the factor structure found in the parent report. Thus, it was determined that the four-factor model identified in the parent-report EFA was the best fitting model for both parent and child reports through the CFA, supporting the hypothesis that the factor structure would be consistent across the parent and youth reports, and providing partial support for the hypothesis that the factor structure in EFA would be replicated in the CFA. The high correlations between the factors on which some items cross-loaded may account for the discrepancies in item loadings between the youth-reported four-factor model in EFA and CFA. Because these scales are highly correlated, it is not surprising that an item may cross-load to varying degrees in different samples. Additionally, a multiple group analysis indicated that the four-factor model met criteria for construct and metric invariance across both parent and youth reports, indicating that this factor structure is consistent for both parent and youth reports. However, scalar invariance was not met for the four-factor model across parent and youth reports which indicates that factor scores between parent and youth reports cannot be directly compared (Steinmetz 2013). In other words, a score of 12 on the parent-report Depression scale is not necessarily equal to a score of 12 on the youth-report Depression scale. However, these scores could be made comparable in the future if they were normed on a representative sample. By collecting Ohio Scales from a large representative sample of both parents and youth, subscale scores could be standardized

into *t*-scores based on the means and standard deviations of the normative sample. This process of standardized score is used in well-established measures of child emotional and behavioral problems (Achenbach and Rescorla 2004). The final four-factor model for the Ohio Scales Problem Severity included the following factors: *Depression*, *Anxiety*, *Aggression*, and *Delinquency*. These subscales are similar to other subscales on well-established measures of child emotional and behavioral problems.

In order to support the concurrent validity of Ohio Scales Problem Severity subscales, these subscales were compared to theoretically similar scales on the CBCL and YSR. Consistent with our hypothesis, the theoretically similar subscales found on the Ohio Scales Problem Severity were highly correlated with the corresponding higher-order factors on the CBCL and YSR. Additionally, dissimilar subscales (e.g., Ohio Scales Problem Severity Delinquency and CBCL Internalizing) had markedly lower correlations, supporting the construct validity of the Problem Severity scale. Given that the CBCL and YSR are considered the gold standard of parent-report and youth-report measures of youth mental health, the high correlations between the Ohio Scales Problem Severity factors and the corresponding factors on CBCL and YSR provide concurrent validity for the subscales of Ohio Scales Problem Severity. Additionally, the high correlation between the corresponding subscales suggest the Ohio Scales Problem Severity scale and CBCL/YSR are measuring similar constructs. While measuring similar constructs, the Ohio Scales Problem Severity scale addresses many of the drawbacks of the CBCL and YSR such as length, complexity of scoring, and cost. Thus, although the CBCL and YSR are the gold standard, our findings suggest that the Ohio Scales Problem Severity scale may provide a more practical measure of child emotional and behavioral problems for CMHs.

The Ohio Scales Problem Severity was designed to be an efficient and cost-effective outcome measure for CMHs, but it lacked valid subscales found in similar well-established but less pragmatic measures. Establishing valid subscales addresses this drawback and increases the clinical utility of the Ohio Scales. Past studies have shown that subscales, such as the Externalizing and Internalizing subscales on the CBCL, have demonstrated greater validity for identifying specific mental health problems compared to an overall problem score (Warnick et al. 2009). Thus, identifying subscales for Ohio Scales Problem Severity scale provides a more specific indicator of behavioral and emotional problems. Having this more specific data regarding clients' needs could improve treatment planning and allow clinicians to better track specific treatment outcomes.

Tracking more specific emotional and behavioral problems in clients will enable clinicians to make more informed treatment decisions at intake and throughout the

treatment process. Studies have shown that different treatments are more effective for specific problems (Chambless and Ollendick 2001; Siev and Chambless 2007; Weisz 2004; Weisz et al. 2006). For example, cognitive therapy is efficacious in changing negative cognitions in depressed adolescents, but it does not necessarily improve externalizing behaviors (Weisz 2004). Thus, examining the subscales on the Ohio Scales Problem Severity scale can provide clinicians with additional information in matching treatment to a youth's specific need. Beyond making the initial treatment plan, the Ohio Scales Problem Severity scale can be used to track outcomes throughout treatment in order to evaluate the changing needs of a client. Tracking outcomes and providing feedback to therapists can lead to better treatment outcomes at discharge because it allows clinician to adjust treatment to the needs of the client over time (Bickman et al. 2011; Goodman et al. 2013; Lambert et al. 2003). For example, Bickman et al. (2011) showed that clinicians who received regular feedback based on a standardized measure of youth symptom severity and functioning had significantly better treatment outcomes than those who did not receive regular feedback. Further, having subscales could provide information about changes in specific symptoms over time, and thus could allow clinicians to make adjustments to treatment to address the changing needs of a client. For example, if a client with comorbid anxiety and depression is reporting improvement in depressive symptom but increased anxious symptoms based on their subscale scores, the clinician may adjust treatment to address symptoms of anxiety. Thus, using the Ohio Scales Problem Severity scale to track outcomes throughout treatment can give clinicians the specific information needed to adjust treatment to meet the changing needs of a client. Additionally, the short and practical nature of the Ohio Scales, compared to longer measures such as the CBCL and YSR, reduces the burden of using them repeatedly throughout treatment.

The current study found a theoretically-sound and empirically-sound factor structure for the Ohio Scales Problem Severity scale, which could provide increased clinical utility both in treatment planning and outcome tracking. The study has many strengths including using archival data from a clinical sample which provided a representative sample of youth who access CMH. Since the Ohio Scales were designed to be utilized within CMH and other child welfare agencies, it is a strength that the data collected for these analyses came from a CMH setting rather than a more controlled research setting. The real world source for these data provides external validity to these findings. Additionally, we had access to multiple reporters and to the CBCL and YRS, the current gold standard for parent and youth report of youth emotional and behavioral problems. Having these additional data provided validity for the subscales

found on the Ohio Scales Problem Severity scale. Finally, due to the size of the sample, we were able to randomly split the sample into the two equivalent subsamples. Using the two subsamples, we were able to use an EFA and then replicate the findings using a CFA.

Despite these strengths, there were limitations to the study. First, although the data provided a clinical sample from a CMH, the CMH was located in a semi-rural, Midwestern county and thus these findings may not generalize to other populations. Second, although it was found that the subscales from the Ohio Scales Problem Severity scale were correlated with the corresponding subscales on the CBCL, this does not address the validity of these scales with corresponding DSM-5 diagnoses. Comparing the Ohio Scales Problem Severity subscales with diagnoses established through structured clinical interviews would strengthen the construct validity of these subscales. Additionally, both the Ohio Scales Problem Severity scale and CBCL/YSR are self-report measures, which may lead to some covariance that can be attributed to method bias. Validating the Ohio Scales Problem Severity subscales with a multi-method approach, such as a structured clinical interview or direct observation, would further validate these subscales. Finally, the data were collected as part of the standard protocol for intake and treatment at a CMH. Given that these data were never intended for research, the data collection process may have been less stringent than if the data had been collected specifically for research. Though the data were cleaned and checked for biases, using real-world data may have introduced additional error due to data entry errors. Replicating our study with data collected in a more controlled environment would strengthen the internal validity of these findings.

As the demand for outcome monitoring in CMHs grows, so does the need for a psychometrically-sound and pragmatic outcome measure. Development and research on measures that meet the needs of CMHs is imperative in improving the quality and cost-efficiency of the public mental health system. The Ohio Scales were specifically designed to meet these needs, and our findings expand the clinical utility of the Ohio Scales Problem Severity scale for CMHs. However, more research is needed to continue to expand the utility and validity of the measure. The link between the Ohio Scales Problem Severity subscales and DSM-5 diagnoses would provide further validity to the subscales. Additionally, longitudinal studies that examined change in the subscale scores over time could show subscales' response to change, which would increase the utility of the Ohio Scales Problem Severity scale as an outcome measure for treatment. Finally, studies using the Ohio Scales Problem Severity scale with a large, nationally-representative sample could be used to establish normative scores for this measure. Any future research that expands

the validity and utility of Ohio Scales will benefit CMHs, which need a psychometrically-sound and practical measure of youth emotional and behavioral problems.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interests.

References

- Achenbach, T. (1991). Manual for the Child Behavior Checklist/4 - 18 and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. A. (2004). The Achenbach System of Empirically Based Assessment (ASEBA) for Ages 1.5 to 18 Years. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Vol. 2. Instruments for children and adolescents* (3rd ed., pp. 179–213). Mahwah, NJ: Lawrence Erlbaum.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.
- Bickman, L., Kelley, S., Breda, C., de Andrade, A., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423–1429. doi:10.1176/appi.ps.002052011.
- Bickman, L., Smith, C. M., Lambert, E. W., & Andrade, A. R. (2003). Evaluation of a congressionally mandated wraparound demonstration. *Journal of Child and Family Studies, 12*(2), 135–156. doi:10.1023/a:1022854614689.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*(1), 685–716. doi:10.1146/annurev.psych.52.1.685.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. doi:10.1207/s15328007sem0902_5.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509. doi:10.1037/0033-2909.131.4.483.
- Dowell, K. A., & Ogles, B. M. (2008). The Ohio Scales youth form: Expansion and validation of a self-report outcome measure for young children. *Journal Of Child And Family Studies, 17*(3), 291–305. doi:10.1007/s10826-007-9137-2.
- Garland, A. F., Haine, R. A., & Boxmeyer, C. L. (2007). Determinates of youth and parent satisfaction in usual care psychotherapy. *Evaluation and Program Planning, 30*(1), 45–54. doi:10.1016/j.evalprogplan.2006.10.003.
- Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services & Research, 30*(4), 393–405. doi:10.1007/bf02287427.
- Goodman, J. D., McKay, J. R., & DePhilippis, D. (2013). Progress monitoring in mental health and addiction treatment: A means of improving care. *Professional Psychology: Research And Practice, 44*(4), 231–246. doi:10.1037/a0032605.
- Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research And Practice, 35*(5), 485–491. doi:10.1037/0735-7028.35.5.485.
- Hayek, M., Mackie, T. I., Mulé, C. M., Bellonci, C., Hyde, J., Bakan, J. S., & Leslie, L. K. (2014). A multi-state study on mental health evaluation for children entering foster care. *Administration and Policy in Mental Health and Mental Health Services Research, 41*(4), 552–567. doi:10.1007/s10488-013-0495-3.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. doi:10.1080/10705519909540118.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Education and Psychological Measurement, 20*, 141–151. doi: 1177/001316446002000116.
- Kasius, M. C., Ferdinand, R. F., van den Berg, H., & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry, 38*(6), 625–632. doi:10.1111/j.1469-7610.1997.tb01689.x.
- Kenny, M. C., & Faust, J. (1997). Mother-child agreement on self-report of anxiety in abused children. *Journal of Anxiety Disorders, 11*(5), 463–472. doi: 10.1016/s0887-6185(97)00023-6.
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science And Practice, 10*(3), 288–301. doi:10.1093/clipsy.bpg025.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. doi:10.1080/01621459.1988.10478722.
- Mataszow, G. S., & Kamphaus, R. W. (2001). Behavior Assessmenteditors System for Children (BASC): Toward accurate diagnosis and effective treatment. In *Handbook of psychoeducational assessment: Ability, achievement, and behavior in children* (pp. 257–288). San Diego, CA: Academic Press.
- Muris, P., Meesters, C., & Berg, Fvanden (2003). The Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry, 12*(1), 1–8. doi:10.1007/s00787-003-0298-2.
- Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (2000). *The Ohio youth problem, functioning, and satisfaction scales: Technical manual*. Columbus, OH: Ohio Department of Mental Health.
- Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (2001). The Ohio Scales: Practical outcome assessment. *Journal Of Child And Family Studies, 10*(2), 199–212. doi:10.1023/a:1016651508801.
- Rosenblatt, A., & Rosenblatt, J. A. (2002). Assessing the effectiveness of care for youth with severe emotional disturbances: Is there agreement between popular outcome measures?. *The Journal of Behavioral Health Services & Research, 29*(3), 259–273. doi:10.1097/00075484-200208000-00002.
- Roth, D. M., & Leavey, G. (2006). Mental health help-seeking and young people: A review. *Pastoral Care in Education, 24*(3), 4–13. doi:10.1111/j.1468-0122.2006.00373.x.
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*(1), 1–10. doi:10.1037/a0018082.
- Siev, J., & Chambless, D. L. (2007). Specificity of treatment effects: Cognitive therapy and relaxation for generalized anxiety and panic disorders. *Journal of Consulting and Clinical Psychology, 75*(4), 513–522. doi:10.1037/0022-006x.75.4.513.
- Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology, 9*(1), 1–12. doi:10.1027/1614-2241/a000049.

- Texas Department of Mental Health and Mental Retardation [TDMHMR] (2004). Validation and norms for the Ohio Scales among children served by the Texas Department of Mental Health and Mental Retardation. *Texas Department of Mental Health and Mental Retardation*. Retrieved 9 May 2014, from <http://www.dshs.state.tx.us/mhprograms/FinalRprtvalidnormsOhioScales.pdf>.
- Trask, E. V., & Garland, A. F. (2011). Are children improving? Results from outcome measurement in a large mental health system. *Administration and Policy in Mental Health and Mental Health Services Research*, 39(3), 210–220. doi:10.1007/s10488-011-0353-0.
- Turchik, J. A., Karpenko, V., & Ogles, B. M. (2007). Further evidence of the utility and validity of a measure of outcome for children and adolescents. *Journal of Emotional and Behavioral Disorders*, 15(2), 119–128. doi:10.1177/10634266070150020201.
- Turchik, J. A., Karpenko, V., Ogles, B. M., Demireva, P., & Probst, D. R. (2010). Parent and adolescent satisfaction with mental health services: Does it relate to youth diagnosis, age, gender, or treatment outcome? *Community Mental Health Journal*, 46(3), 282–288. doi:10.1007/s10597-010-9293-5.
- U.S. Census Bureau (2014, June 11). State & County Quickfacts: Wood County, OH. Retrieved 22 June 2014, from <http://quickfacts.census.gov/qfd/states/39/39173.html>.
- Warnick, E. M., Weersing, V. R., Scahill, L., & Woolston, J. L. (2009). Selecting measures for use in child mental health services: A scorecard approach. *Administration and Policy in Mental Health and Mental Health Services Research*, 36(2), 112–122. doi:10.1007/s10488-008-0203-x.
- Weisz, J. R. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics of North America*, 13(4), 729–815. doi:10.1016/j.chc.2004.05.006.
- Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin*, 132(1), 132–149. doi:10.1037/0033-2909.132.1.132.