
Reliability and Validity of the Youth Outcome Questionnaire Self-Report



Nathanael W. Ridge, Jared S. Warren, Gary M. Burlingame,
M. Gawain Wells, and Katherine M. Tumblin

Brigham Young University

The psychometric properties of an adolescent mental health outcome instrument (Youth Outcome Questionnaire Self-Report version [Y-OQ-SR]) were examined. Participants were 206 adolescents (ages 12–18; mean age = 15). The Y-OQ-SR was evaluated in terms of its internal consistency, test-retest reliability, and concurrent validity. Criterion measures included the Behavior Assessment System for Children, Self-Report of Personality—Adolescent Version (SRP), and the Child Behavior Checklist Youth Self Report (YSR). Analyses revealed very good internal consistency and test-retest reliability of the Y-OQ-SR total score and subscales, and moderate to good concurrent validity with the SRP and YSR. The Y-OQ-SR appears to be a valid and reliable self-report measure of psychosocial distress that warrants further study in youth psychotherapy research. © 2009 Wiley Periodicals, Inc. *J Clin Psychol* 65: 1115–1126, 2009.

Keywords: youth; treatment outcome; self-report; psychometric properties; psychotherapy

Rising health care costs have resulted in an increased focus on accountability from policymakers, administrators, service providers, and consumers of mental health services (Krupnick & Pincus, 1992). The percentage of the gross domestic product (GDP) of the United States used on health care is estimated to climb to 17% of the GDP by 2011 (Maruish, 2004), and a notable portion of these health care costs will be used for mental health services (Harwood et al., 2003; Mark et al., 2005). As a result, mental health care has entered the “era of accountability,” (Burlingame, Lambert, Reisinger, Neff, & Mosier, 1995) where clinicians are expected to systematically evaluate the outcomes of clinical interventions.

Outcome measures are valuable tools to demonstrate patient improvement in efficacy and effectiveness research, and, more recently, patient-focused research. The continued refinement of outcome assessments has allowed for reliable conclusions in

Correspondence concerning this article should be addressed to: Jared S. Warren, Department of Psychology, Brigham Young University, 291 TLRB, Provo, UT 84602; e-mail: jared_warren@byu.edu

treatment efficacy and effectiveness research, where patients' responses to treatment are measured in highly controlled experimental or more naturalistic conditions, respectively (Kazdin, 1994; Lambert & Hawkins, 2004). As a complementary paradigm to the traditional pre, post and follow-up method of outcome assessment, Howard, Moras, Brill, Martinovich, and Lutz (1996) introduced a third approach described as "patient-focused research," which is aimed at monitoring an individual patient's progress over the course of therapy with multiple data points (Lambert, Hansen, & Finch, 2003). Patient-focused outcome monitoring answers the question, "Is this treatment working for this patient at this time?" (Burlingame et al., 2001) and assists in continuous quality improvement (CQI), which has become the required standard of accrediting and managed care organizations (Maruish, 2004). Unfortunately, few outcome measures exist that are appropriate for use on a continual session-by-session basis (Froyd, Lambert, & Froyd, 1996).

Of the few psychometrically sound outcome measures used to demonstrate efficacy and effectiveness of mental health treatment for children and adolescents, many have attributes that preclude their use on a session-by-session basis. For instance, commonly used measures like the Behavior Assessment System for Children (Reynolds & Kamphaus, 2004) or the Child Behavior Checklist (Achenbach, 1999), despite excellent psychometric properties, have been criticized as too lengthy, too expensive, and not sensitive enough to change to be administered on a continual or session-by-session basis (Burns, 2002; Drotar, Stein, & Perrin, 1995; Perrin, Stein, & Drotar, 1991).

In contrast to these measures, the Youth Outcome Questionnaire (Y-OQ; Wells, Burlingame, & Rose, 2003) was designed to address many of the limitations of other measures in regularly monitoring client outcomes. The Y-OQ is a parent-report measure developed to facilitate the work of researchers, managed care companies, and clinicians in tracking outcome on a continual, session-by-session basis. In addition, the Y-OQ was "constructed to be brief, sensitive to change over short periods of time, and available at a nominal cost while maintaining high psychometric standards of reliability and validity" (Burlingame, Wells, Lambert, & Cox, 2004, p. 238).

Although the parent-report version of the Y-OQ has been demonstrated to be a cost-effective tool that can be used as a CQI instrument, parents and youth typically show weak agreement in their ratings of youth symptomatology, which suggests that "no single source of data can serve as a gold standard" for youth symptoms (Achenbach, McConaughy, & Howell, 1987; Achenbach, 1999, p. 429). Parents are not always aware of deviant adolescent behavior outside of the home and are often poor reporters of adolescent internalizing behaviors such as symptoms of depression and anxiety (Comer & Kendall, 2004; Meiser-Stedman, Smith, Glucksman, Yule, & Dalgleish, 2007). Consequently, multiple informants are needed to effectively measure mental health outcomes in youth.

The Youth Outcome Questionnaire Self-Report (Y-OQ-SR) was designed to be a reliable and change-sensitive self-report measure of psychosocial distress in adolescents that can be used on a session-by-session basis. The Y-OQ-SR was developed with the same goals as the original Y-OQ; however, the Y-OQ-SR allows measurement of behavior change, as an adolescent perceives it (Wells et al., 2003). Pilot research has yielded preliminary support for the Y-OQ-SR as a psychometrically sound measure of adolescent treatment outcomes; however, conclusions have been limited by small samples and the need to demonstrate concurrent validity with

existing measures of youth behavior problems (Wells et al., 2003). The purpose of the present study was to examine the Y-OQ-SR's reliability through internal consistency estimates and test-retest methods, and to examine its concurrent validity by investigating its relationship with established instruments of adolescent psychopathology (i.e., BASC-2 Self Report of Personality Adolescent Version & CBCL Youth Self Report).

Method

Participants

This study reports the results obtained from a non-clinical community sample of 206 youth ages 12 to 18 (129 females, 77 males; mean age = 14.72, $SD = 1.87$) recruited from an urban community in the intermountain west of the United States. The average grade for the community sample was the 9th grade ($SD = 1.89$). Racial identity was self-identified as follows: Caucasian ($n = 179$; 92%), Hispanic/Latin American ($n = 6$; 3%), Asian American ($n = 3$; 2%), and Other ($n = 7$; 3%).

Measures

Youth Outcome Questionnaire—Self-Report Version 2.0 (Y-OQ-SR; Wells et al., 2003; see also Burlingame, Wells, Lambert, & Cox, 2004). The Y-OQ-SR is a self-report measure of treatment progress for adolescents ages 12 to 18. The Y-OQ-SR comprises 64 items with six subscales designed to assess several behavioral domains of children and adolescents experiencing behavioral difficulties. The Y-OQ-SR yields a total distress score plus subscale scores for the following domains: intrapersonal distress (e.g., anxiety, depression, hopelessness; 18 items); somatic complaints (e.g., headaches, dizziness, stomachaches; 8 items); interpersonal relations (e.g., arguing, defiance, communication problems; 10 items); social problems (e.g., delinquent or aggressive behaviors; 8 items); behavioral dysfunction (e.g., organization, concentration, handling frustration, and ADHD-related symptoms; 11 items); and critical items (symptoms often found in youth receiving inpatient services, such as paranoid ideation, hallucinations, mania, and suicidal feelings; 9 items). The measure takes approximately 7 minutes to complete. Items are presented in a 5-point scale with options including 0 (*never*), 1 (*rarely*), 2 (*sometimes*), 3 (*frequently*), and 4 (*almost always*). Seven of the items are written and reverse-scored to describe elements of healthy behavior and are weighted differently, with scores ranging from 2 to -2 . Norms for community and clinical samples (including outpatient, residential, and partial hospitalization) are provided by the test authors (Wells et al., 2003).

Preliminary pilot studies (Wells et al., 2003) indicated that the Y-OQ-SR is highly reliable, reporting an internal consistency estimate of .96 across both clinical and community samples ($n = 1334$) for the total distress score. Reliability estimates for the subscales were as follows: intrapersonal distress (.91), somatic complaints (.73), interpersonal relations (.77), social problems (.84), behavior dysfunction (.78), and critical items (.81).

Behavior Assessment System for Children-2—Self-Report of Personality (BASC-2 SRP-A; Reynolds & Kamphaus, 2004). The SRP is a personality inventory used to measure the emotions and self-perceptions of adolescents between the ages of 12 to 18. This measure contains 176 items and comprises 16 subscales, with five composite

scales. The measure takes approximately 20 to 30 minutes to complete. Item response formats include true = 1 or false = 0, or a 4-point scale: 0 (*never*), 1 (*sometimes*), 2 (*often*) and 3 (*almost always*).

Reynolds and Kamphaus (2004) reported that the SRP-A has demonstrated good reliability. Internal consistency estimates range from .67 to .88 for the subscales and .84 to .95 for the composite scores. Test-retest reliability estimates range from .63 to .84 for the subscales and .76 to .84 for the composite scores. Reynolds and Kamphaus (2004; Kamphaus, Reynolds, Hatcher, & Kim, 2004) also report that the SRP-A has demonstrated good validity.

Child Behavior Checklist Youth Self Report (CBCL-YSR; Achenbach, 1999).

The YSR is a self-report measure of behavioral and emotional problems in youth ages 11 to 18. This measure contains 113 items and comprises nine subscales and two composite scores. It takes approximately 15 to 20 minutes to complete the behavioral/emotional problems section of the YSR. Adolescents rate themselves on a scale from 0 to 2 on each item (0 = not true, 1 = *somewhat true or sometimes true*, and 2 = *very true or often true*).

Achenbach (1999) reported good reliability for the YSR. Internal consistency estimates range from .59 to .86 for the subscales, while estimates for internalizing, externalizing, and total score are .89, .89, and .95, respectively. Test-retest reliability estimates are reported from .57 to .81 for the subscales at 1 week, while estimates for internalizing, externalizing, and total score at one week are .80, .81, and .79, respectively. Achenbach (1999; Achenbach & Rescorla, 2004) also report that the YSR has demonstrated good validity.

Procedure

Youth ages 12 to 18 were recruited for participation in one of two ways. In the first group, 350 research packets were given to youth at schools, which included the aforementioned questionnaires, age-appropriate information questionnaires, a cover sheet, consent and assent forms, and a self-addressed, prepaid envelope. A small monetary donation to their respective schools was offered as an incentive. Classes were selected based on their availability for recruitment. Sixty-one youth returned the questionnaires through the mail, yielding a return rate of 17.4%, and usable data were obtained for 58 participants. Follow-up questionnaires were sent to each participant, and 27 responded, yielding a return rate of 46.6% for the second mailing. The average time between test administrations was 27 days ($SD = 11.78$).

In the second group, 7,055 residents from a large western community with a population of approximately 443,738 (U.S. Census Bureau, 2007) were randomly telephoned. A group of 18 undergraduate telephone solicitors inquired if youth within the age range of 12 to 18 and not currently receiving mental health treatment were living in the home. If such youth were present, parents or guardians were asked to allow a specified youth to complete a research packet. Four hundred and twenty-five research packets were mailed out with both consent and assent forms included. One hundred and forty-five youth returned the questionnaires, yielding a return rate of 34.1%, and usable data were obtained for 143 participants. Follow-up questionnaires were sent to each of these individuals; 89 responded, yielding a return rate of 62.2% for the second mailing.

Data Analysis

Statistical analyses were calculated using SPSS 16.0 for Windows. Consistent with author recommendations (Wells et al., 2003), cases with more than five items missing were excluded ($n = 5$), yielding a community sample of 201. Internal consistency was calculated for the all subscales and total distress score of the Y-OQ-SR; for test-retest reliability, correlations were calculated between the first and second administrations of the Y-OQ-SR. To determine concurrent validity of the Y-OQ-SR, subscale raw totals were calculated for the Y-OQ-SR, CBCL YSR, and BASC-2 SRP-A, and Pearson Product Moment Correlations were calculated. Convergent validity was assessed by noting correlations between the Y-OQ-SR subscales and conceptually similar criterion subscales measured by the BASC-2 SRP-A and CBCL YSR. Divergent validity was assessed by using a t test for non-independent correlations (Steiger, 1980) among the Y-OQ-SR subscales and the predicted criterion subscales compared with the correlations with the non-criterion subscales.

Results

Descriptive Statistics

The frequency distributions for the Y-OQ-SR were minimally positively skewed and leptokurtic. Significant differences were observed for chi-square analyses of race, $\chi^2(5) = 26.503, p < .001$. In this community sample ($n = 201$), total scores averaged 34.37 ($SD = 29.42$), ranging from -16 to 151 ($SD = 37.00$; negative total scores on the Y-OQ-SR can occur due to items measuring adaptive behaviors that are reverse scored).

Internal Consistency Reliability

The Y-OQ-SR demonstrated moderate to high internal consistency reliability. Alpha coefficients for the Y-OQ-SR subscales were as follows: intrapersonal distress (.91), behavioral dysfunction (.81), somatic complaints (.78), interpersonal relations (.75), social problems (.71), and critical items (.74). Coefficient alpha for the total score was .95.

Test-Retest Reliability

Of the 206 participants who returned an initial, valid Y-OQ-SR, 117 participants returned a second Y-OQ-SR (mean test-retest interval = 27 days); however, six individuals reported experiencing a traumatic event, taking psychotropic drugs or beginning mental health therapy between test administrations. Their questionnaires were omitted, yielding a sample of 111 for the analysis. No significant difference in initial Y-OQ-SR total scores was found between those who completed a second Y-OQ-SR and those who did not ($t = 1.539, p = .127$). Overall, the results indicated high test-retest correlations for the total score ($r = .89, p < .001$) and moderate to high test-retest correlations for individual subscale scores, ranging from .68 to .86 (see Table 1).

Concurrent Validity

As hypothesized, virtually all of the correlations between the Y-OQ-SR and conceptually similar criteria on both the CBCL YSR and the BASC SRP reached

Table 1
 Test-Retest Reliability Coefficients for Community Sample ($n = 111$)

Scale	Time 1		Time 2		Pearson r
	Mean	SD	Mean	SD	
Intrapersonal distress	13.97	10.89	11.4	11.78	.86**
Behavioral dysfunction	8.66	6.41	7.26	6.95	.81**
Somatic complaints	5.37	4.09	4.72	4.12	.83**
Interpersonal relations	-0.27	4.34	-0.27	4.97	.78**
Social problems	0.23	3.07	-0.24	2.97	.68**
Critical items	4.54	3.84	3.83	3.68	.79**
Total	32.51	27.51	26.69	30.13	.89**

Note. SD indicates standard deviation. ** $p < .001$ level.

Table 2
 Intercorrelations Between the Y-OQ-SR and YSR for Community Sample ($n = 201$)

	Y-OQ-SR subscales						Total
	SC	ID	BD	SP	CI	IR	
<i>YSR subscales</i>							
Withdrawn	.40**	.66**	.40**	.28**	.52**	.46**	.58**
Somatic	.68**	.55**	.49**	.41**	.54**	.47**	.62**
Anxiety/depression	.54**	.81**	.56**	.45**	.73**	.57**	.76**
Social problems	.37**	.54**	.49**	.27**	.47**	.44**	.54**
Thought problems	.44**	.51**	.57**	.42**	.64**	.47**	.61**
Attention problems	.52**	.62**	.75**	.49**	.65**	.58**	.72**
Delinquent behavior	.39**	.52**	.51**	.69**	.55**	.64**	.63**
Aggressive	.42**	.50**	.66**	.49**	.54**	.61**	.64**
Internalizing	.62**	.80**	.58**	.46**	.72**	.59**	.78**
Externalizing	.46**	.57**	.68**	.62**	.61**	.69**	.71**
YSR Total	.61**	.78**	.69**	.58**	.75**	.70**	.83**

Note. Y-OQ-SR indicates Youth Outcome Questionnaire Self-Report; YSR, Youth Self Report; ID, interpersonal distress; BD, behavioral dysfunction; SC, somatic complaints; IR, interpersonal relations; SP, social problems; CI, critical items. ** $p < .001$ level.

statistical significance (see Tables 2 and 3). The highest correlation was between the Y-OQ-SR total score and the CBCL YSR total problems scale ($r = .83$), while correlations between similar constructs ranged between .61 and .81. The pattern of moderate to high subscale relationships suggests convergent validity among subscales on the Y-OQ-SR, the CBCL YSR, and BASC-2 SRP-A.

Several exceptions were found. First, the critical items subscale correlated moderately with nearly all criterion subscales on both measures. However, this subscale was not meant to measure a single underlying construct. Rather, this scale contains critical screening items that are common across many constructs; therefore, moderate correlations across criterion subscales would be expected. Also of note, the social problems and interpersonal relations subscales of the Y-OQ-SR failed to correlate highly with BASC-2 SRP-A subscales because of limited externalizing constructs (e.g., aggression or oppositional behaviors) found on the SRP-A rather

Table 3
Intercorrelations Between the Y-OQ-SR and BASC-2 SRP-A Subscales for Community Sample (n = 201)

	Y-OQ-SR subscales						
	SC	ID	BD	SP	CI	IR	TOT
<i>BASC-2 SRP-A subscales</i>							
AS	.37**	.56**	.54**	.40**	.49**	.51**	.59**
AT	.35**	.54**	.52**	.44**	.46**	.57**	.58**
SS	.07	.03	.23**	.22**	.12	.17*	.15*
ScP	.35**	.50**	.58**	.46**	.47**	.56**	.58**
Atyp	.43**	.57**	.51**	.45**	.66**	.51**	.62**
LoC	.36**	.58**	.52**	.46**	.58**	.55**	.62**
SocS	.47**	.73**	.53**	.43**	.56**	.58**	.69**
A	.52**	.70**	.49**	.29**	.64**	.40**	.65**
D	.46**	.78**	.56**	.51**	.61**	.62**	.74**
SoI	.44**	.67**	.65**	.55**	.64**	.60**	.72**
Som	.67**	.48**	.46**	.34**	.52**	.42**	.57**
Intern	.58**	.82**	.66**	.54**	.76**	.65**	.82**
Atte	.46**	.55**	.72**	.52**	.53**	.62**	.68**
Hyp	.29**	.33**	.61**	.33**	.34**	.42**	.46**
In/Hy	.43**	.51**	.75**	.50**	.51**	.60**	.65**
RwP	-.27**	-.52**	-.52**	-.58**	-.52**	-.62**	-.60**
InR	-.44**	-.64**	-.42**	-.34**	-.50**	-.52**	-.60**
SE	-.50**	-.71**	-.46**	-.43**	-.57**	-.51**	-.66**
SR	-.26**	-.48**	-.45**	-.33**	-.39**	-.44**	-.49**
PAC	-.45**	-.75**	-.60**	-.56**	-.65**	-.68**	-.75**

Note. Y-OQ-SR indicates Youth Outcome Questionnaire Self-Report; BASC-2 SRP-A, Behavior Assessment System for Children-2—Self-Report of Personality; ID, interpersonal distress; BD, behavioral dysfunction; SC, somatic complaints; IR, interpersonal relations; SP, social problems; CI, critical items; TOT, total score for Y-OQ-SR; AS, attitude to school; AT, attitude to teachers; SS, sensation seeking; ScP, school problems composite; Atyp, atypicality; LoC, locus of control; SocS, social stress; A, anxiety; D, depression; SoI, sense of inadequacy; Som, somatization; Intern, internalizing problems composite; Atte, attention problems; Hyp, hyperactivity; In/Hyp, inattention/hyperactivity composite; RwP, relations with parents; InR, interpersonal relations; SE, self-esteem; SR, self-reliance; PAC, personal adjustment composite. * $p < .01$ level, ** $p < .001$ level.

than poor convergent validity; these were excluded from subsequent divergent validity analyses.

Moderate support was found for the divergent validity of the Y-OQ-SR subscales (see Tables 4 and 5). Initially, this was assessed by visually noting smaller correlations between the dissimilar criterion subscales and larger correlations between similar criterion subscales with the Y-OQ-SR. Statistically, t tests between conceptually similar criterion and Y-OQ-SR subscales as compared with dissimilar criterion yielded mixed evidence. For instance, when the Y-OQ-SR subscales were compared with the CBCL YSR subscales, excellent divergent validity was found for the somatic complaints, intrapersonal distress, and social problems subscales, while the behavioral dysfunction and interpersonal relations subscales had one or more dissimilar criterion correlations that were not statistically different from the predicted similar criterion correlations. When compared with the BASC-2 SRP-A, both the intrapersonal distress and the behavioral dysfunction subscales had two similar criteria that yielded moderate to high convergent validity estimates; however, on some t tests, only the more highly correlated criterion was significantly different from dissimilar subscales' correlations.

Table 4

Validity Estimates Between Y-OQ-SR Subscales and CBCL-YSR Subscales ($n = 201$)

	Somatic complaints	Intrapersonal distress	Behavioral dysfunction	Social problems	Interpersonal relations
Withdrawn	.40*	.66*	.40*	.28*	.46
Somatic	<u>.68</u>	.55*	.49*	.41†	.47
Anxiety/ depression	.54*	<u>.81</u>	.56*	.45†	.57
Social problems	.37*	.54*	.49*	.27*	.44*
Thought problems	.44*	.51*	.57*	.42†	.47*
Attention problems	.52*	.62*	<u>.75</u>	.49†	.58*
Delinquent Aggressive	.39* .42*	.52* .50*	.51* .66	.69 .49	.64* <u>.61</u>

Note. Y-OQ-SR indicates Youth Outcome Questionnaire Self-Report; CBCL-YSR, Child Behavior Checklist-Youth Self Report. Underlined value corresponds to hypothesized "criterion" index and should have highest overall value.

*Indicates significant value ($p < 0.01$) on t test comparison of differences among all hypothesized criterion and divergent subscales in the same column.

†Indicates significant value ($p < 0.01$) on t -test comparison of differences between one of two hypothesized criterion and divergent subscales in the same column.

Table 5

Validity Estimates Between Y-OQ-SR Subscales and BASC-2 SRP-A Subscales ($n = 201$)

Scale	Somatic complaints	Intrapersonal distress	Behavioral dysfunction
Attitude to school	.37*	.56*	.54†
Attitude to teachers	.35*	.54*	.52
Sensation seeking	.07*	.03*	.23*
Atypicality	.43*	.57*	.51†
Locus of control	.36*	.58†	.52†
Social stress	.47*	.73	.53†
Anxiety	.52*	.70	.49†
Depression	.46*	.78	.56†
Sense of inadequacy	.44*	.67†	.65
Somatization	.67	.48*	.46†
Attention problems	.46*	.55*	.72
Hyperactivity	.29*	.33*	.61

Note. Y-OQ-SR indicates Youth Outcome Questionnaire Self-Report; BASC-2 SRP-A, Behavior Assessment System for Children-2—Self-Report of Personality. Underlined value corresponds to hypothesized criterion index and should have highest overall value.

*Indicates significant value ($p < 0.01$) on t test comparison of differences between all hypothesized criterion and divergent subscales in the same column.

†Indicates significant value ($p < 0.01$) on t test comparison of differences between one of two hypothesized criterion and divergent subscales in the same column.

Discussion

Reliability

Consistent with preliminary estimates (Wells et al., 2003), the Y-OQ-SR total score demonstrated high internal consistency, suggesting the value of using the total score as a measure of severity of psychological distress (Burlingame et al., 2001). Subscale

internal consistency estimates ranged from .71 to .91, indicating that the subscale domains possess moderate to high internal consistency for the hypothesized constructs.

This study found that the Y-OQ-SR's total score and subscale scores possess moderately high temporal stability with test-retest reliability coefficients between .68 and .86 for the subscales and .89 for the total score (Burlingame et al., 2001). Overall, the test-retest coefficient estimates met or exceeded .70, the general recommendation for outcome measure, test-retest reliability (Burlingame et al., 1995). Of note, the intrapersonal distress subscale demonstrated remarkably high temporal stability and internal consistency relative to the total score. These high estimates, in contrast to some research on adolescent self-report data (Hartung, McCarthy, Milich, & Martin, 2005), suggest that adolescents can be reliable reporters of their psychological distress. The Y-OQ-SR appears to provide reliable ratings for constructs (e.g., intrapersonal distress) inaccessible by other raters (Cantwell et al., 1997; Derogatis & Culpespper, 2004). Furthermore, the temporal stability of the Y-OQ-SR over an average of a 4-week period indicates changes in scores can be considered reflective of change in symptom severity rather than random fluctuations (Burlingame et al., 2001).

Concurrent Validity

The Y-OQ-SR demonstrated strong relationships with other frequently used self-report measures of adolescent psychopathology (i.e., BASC-2 SRP-A, CBCL YSR). Y-OQ-SR subscales converged with corresponding theoretical constructs measured by the SRP-A and YSR. More specifically, these intercorrelations achieved the minimum cutoff of .5 and exceeded the criterion of .75, indicating excellent validity (Burlingame et al., 1995, p. 229). Moderate support was found for the divergent validity of the Y-OQ-SR subscales. Specifically, the absolute values between dissimilar criterion subscales and larger correlations between similar criterion subscales with the Y-OQ-SR were noted. Thirty-six of the forty comparisons between unrelated subscales of the CBCL YSR and convergent Y-OQ-SR criterion correlations were significantly lower than correlations between predicted criterion of the YSR and convergent Y-OQ-SR subscales. Similar results were found when the BASC-2 SRP-A subscales were used (47 of 62). These results lend preliminary support for the utility of the individual Y-OQ-SR subscales; however, interpretations of subscale scores should be made with caution until factor analytic evidence supports their value as empirically distinct domains.

Study Limitations

Several limitations should be considered in evaluating the methods and results of this study. First, sampling bias may limit the generalizability of findings. The first subset of adolescents recruited from their respective schools was not randomly selected. In addition, although chi-square analyses of demographics between the randomized sample and convenience sample revealed little to no differences, sampling bias could apply to the randomized sample as only students and parents that were willing to complete the questionnaires were included. Consequently, the generalizability of the results should be limited to adolescents that demographically reflect these samples. Finally, although this study examined concurrent validity of the Y-OQ-SR with two other widely-used youth self-report measures, additional support for the measure's validity could have been gained through additional criterion measures such as reports from parents or teachers, structured interview data, or direct observations of youth

behavior. Procedures using multiple informants and multiple methods are likely to yield the most accurate information on youth mental health status and outcomes.

Implications for Research, Policy, and Practice

In considering recommendations for evaluating the psychometric rigor of patient-focused outcome instruments (Vermillion & Pfeiffer, 1993), further research is needed to determine the Y-OQ-SR's sensitivity to therapeutic change and evaluate the psychometric properties of the measure with a more demographically and clinically diverse sample. Future research should include youth from multiple treatment settings to understand the potential utility of each item in different treatment contexts and to increase the generalizability to other adolescent populations. Similarly, it would be valuable to determine whether measures such as the Y-OQ-SR could be accurately used as a screening tool for assigning cases to appropriate levels of intervention (e.g., outpatient, inpatient, day treatment).

Although questions remain, the results from this study indicated the Y-OQ-SR possesses acceptable levels of internal consistency, test-retest reliability, and concurrent validity and support the use of the Y-OQ-SR as a global measure of psychosocial distress for adolescent self-reports. In sum, the Y-OQ-SR is a psychometrically sound instrument by standards for patient-focused outcome instruments (Burlingame et al., 2001; Ficken, 1995; Newman, Rugh, & Ciarlo, 2004; Vermillion & Pfeiffer, 1993). Future studies should explore the measure's sensitivity to change, expand its normative sample to increase its generalizability, and examine its usefulness as a screening instrument.

References

- Achenbach, T.M. (1999). The child behavior checklist and related instruments. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (2nd ed., pp. 429–466). Mahwah, NJ: Lawrence Erlbaum Associates.
- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213–232.
- 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 outcome assessment* (3rd ed., Vol. 2, pp. 179–214). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burlingame, G.M., Lambert, M.J., Reisinger, C.W., Neff, W.L., & Mosier, J.I. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration*, 22, 226–236.
- Burlingame, G.M., Mosier, J.I., Wells, M.G., Atkin, Q.G., Lambert, M.J., Whoolery, M., et al. (2001). Tracking the influence of mental health treatment: The development of the youth outcome questionnaire. *Clinical Psychology and Psychotherapy*, 8, 361–379.
- Burlingame, G.M., Wells, M.G., Lambert, M.J., & Cox, J.C. (2004). Youth outcome questionnaire (Y-OQ). In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 2, pp. 235–274). Mahwah, NJ: Lawrence Erlbaum Associates.
- Burns, M.K. (2002). Self-report objective measures of personality for children: A review of psychometric properties for RQC. *Psychology in the Schools* 39, 221–246.

- Cantwell, D.P., Lewinsohn, P.M., Rohde, P., & Seeley, J.R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 610–619.
- Comer, J.S., & Kendall, P.C. (2004). A symptom-level examination of parent-child agreement in the diagnosis of anxious disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 37(7), 878–886.
- Derogatis, L.R., & Culppepper, W.J. (2004). Screening for psychiatric disorders. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 1, pp. 65–110). Mahwah, NJ: Lawrence Erlbaum Associates.
- Drotar, D., Stein, R.E.K., & Perrin, E.K. (1995). Methodological issues in using the child behavior checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology*, 24, 184–192.
- Ficken, J. (1995). New directions for psychological testing. *Behavioral Health Management*, 20, 12–14.
- Froyd, J.E., Lambert, M.J., & Froyd, J.D. (1996). A review of practices of psychotherapy outcomes measurement. *Journal of Mental Health*, 5, 11–15.
- Hartung, C.M., McCarthy, D.M., Milich, R., & Martin, C.A. (2005). Parent-adolescent agreement on disruptive behavior symptoms: A multitrait-multimethod model. *Journal of Psychopathology and Behavioral Assessment*, 27, 159–168.
- Harwood, H.J., Mark, T.L., McKusick, D.R., Coffey, R.M., King, E.C., & Genuardi, J.S. (2003). National spending on mental health and substance abuse treatment by age of clients: 1997. *Journal of Behavioral Health Services & Research*, 30, 433–443.
- Howard, K.I., Moras, K., Brill, P.L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51, 1059–1064.
- Kamphaus, R.W., Reynolds, C.R., Hatcher, N.M., & Kim, S. (2004). Treatment planning and evaluation with the behavior assessment system for children (BASC). In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 2, pp. 331–354). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kazdin, A.E. (1994). Psychotherapy for children and adolescents. In A.E. Bergin & S.L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 543–594). New York: John Wiley.
- Krupnick, J.L., & Pincus, H.A. (1992). The cost-effectiveness of psychotherapy: A plan for research. *American Journal of Psychiatry*, 149, 1295–1305.
- Lambert, M.J., Hansen, N.B., & Finch, A.E. (2003). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting & Clinical Psychology*, 69, 159–172.
- Lambert, M.J., & Hawkins, E.J. (2004). Use of psychological tests for assessing treatment outcomes. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 1, pp. 171–196). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mark, T.L., Coffey, R.M., Vandivort-Warren, R., Harwood, H.J., King, E.C., & the MHSA Spending Estimates Team. (2005). U.S. spending for mental health and substance abuse treatment, 1991–2001. *Health Affairs*, 24, w133–w142 (published online 29 March 2005; 10.1377/hlthaff.w5.133).
- Maruish, M.E. (2004). Introduction. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed., Vol. 1, pp. 1–64). Mahwah, NJ: Lawrence Erlbaum Associates.
- Meiser-Stedman, R., Smith, P., Glucksman, E., Yule, W., & Dalgleish, T. (2007). Parent and child agreement for acute stress disorder, post-traumatic stress disorder and other psychopathology in a prospective study of children and adolescents exposed to single-event trauma. *Journal of Abnormal Child Psychology*, 35(2), 191–201.

- Newman, F.L., Rugh, D., & Ciarlo, J.A. (2004). Guidelines for selecting psychological instruments for treatment planning and outcomes assessment. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed., Vol. 1, pp. 197–214). Mahwah, NJ: Lawrence Erlbaum Associates.
- Perrin, E.C., Stein, R.E.K., & Drotar, D. (1991). Cautions in using the child behavior checklist: Observations based on research about children with chronic illness. *Journal of Pediatric Psychology, 16*, 411–421.
- Reynolds, C.R., & Kamphaus, R.W. (2004). *Behavior assessment for children manual* (2nd ed.). Circle Pines, MN: American Guidance Service, Inc.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.
- U.S. Census Bureau. (2007). Utah County, Utah. Retrieved March 6, 2007 from <http://quickfacts.census.gov/qfd/states/49/49049.html>.
- Vermillion, J.M., & Pfeiffer, S. (1993). Treatment outcome and continuous quality improvement: Two aspects of program evaluation. *Psychiatric Hospital, 24*, 9–14.
- Wells, M.G., Burlingame, G.M., & Rose, P.M. (2003). *Youth outcome questionnaire self report*. Wilmington, DE: American Professional Credentialing Services.