

Giving Youth a Voice: A Preliminary Study of the Reliability and Validity of a Brief Outcome Measure for Children, Adolescents, and Caretakers*

Barry L. Duncan, Psy.D.

*Institute for the Study of Therapeutic Change
Ft. Lauderdale, Florida*

Jacqueline A. Sparks, Ph.D.

*University of Rhode Island
Kingston, Rhode Island*

Scott D. Miller, Ph.D.

*Institute for the Study of Therapeutic Change
Chicago, Illinois*

Robert T. Bohanske, Ph.D.

*Southwest Behavioral Health Services
Phoenix, Arizona*

David A. Claud, MS

*The Center for Family Services of Palm Beach County
West Palm Beach, Florida*

Policy makers and payers are insisting that to be paid, therapists must “deliver the goods.” Concurrently, there is a worldwide movement to involve consumers in their care. Consequently, the measurement of change, from the client’s perspective, has become an important topic. Unfortunately, no self-report outcome measure has been available for children under 13. This article describes the development, validation, and psychometric properties of a brief outcome measure, the Child Outcome Rating Scale (CORS), for use with children 6-12 and their caretakers as well as the Outcome Rating Scale (ORS), for youth 13 and above and their caretakers. Results indicate that the CORS/ORS represent a balanced trade-off between the reliability and validity of longer measures, and the feasibility of these brief scales, and that youth do indeed deserve a voice in the delivery of services. We argue that outcome management is a pivotal part of delivering consumer-driven services, the therapeutic relationship, and change itself.

* The authors wish to thank Jeb Brown for his invaluable statistical consultation.

Requests for offprints should be sent to Dr. Sparks at University of Rhode Island, Transition Center, 2 Lower College Rd., Kingston, RI 02881 or jsparks@uri.edu

More than any previous time in the history of the field, policy makers and payers are stridently insisting that to be paid, therapists, and the systems of care in which they operate, must “deliver the goods.” Accountability is the watchword of the day and “return on investment” the guiding metric. Like it or not, mental health and substance abuse services have become a commodity. Those footing the bill want proof of the effectiveness and value of the product being purchased (Miller, Duncan, Sorrell, & Brown, 2005).

Concurrently there is a growing worldwide movement, both private and governmental, to involve consumers in mental health and substance abuse care (Bohanske, in press). Consequently, the measurement and management of change, from the client’s perspective, has become an important topic in the delivery of mental health services (Brown, Burlingame, Lambert, Jones, & Vaccaro, 2001; Duncan, Miller, & Sparks, 2004), and for good reason: Monitoring client-based outcome, when combined with feedback to the clinician, increases the effectiveness of clinical services by an amazing 65% in real clinical settings (Lambert et al., 2003; Miller, Duncan, Brown, Sorrell, & Chalk, 2006).

Unfortunately, these substantial benefits of outcome management and the move toward consumer-driven services have not included children; there are no valid self-report outcome tools for children under the age of 13. Although numerous child assessment instruments are available, few measure the outcome or benefit of provided services—most were developed to assess psychosocial dysfunction or psychological disorders (Burlingame, Mosier, et al., 2001). Moreover, even if designed for outcome assessment, available measures for children are lengthy and perhaps too cumbersome to be feasible—for both clients and clinicians—for everyday use. Finally, all available outcome measures for children, like the Youth Outcome Questionnaire 30 (YOQ) (Burlingame, Jasper, et al., 2001) are parent-report measures. Adolescents can report on their progress, but children under 13 have no such opportunity. Indeed, background research conducted for this study revealed an astonishing lack of references for not only child self-report outcome measures but also about children’s perceptions about therapy in general (Day, Carey, & Surgenor, 2006).

In short, children have little voice in the services they receive. More often than not, children and adolescents receive services as mandated clients—someone else thought they needed it. This often leaves the youth with little control over the process, and perhaps, little reason to engage. It is no wonder, then, that 40-60% of youth drop out of treatment (Kasdin, 2004). To fill the void regarding self-report measures for the under 13 age group, address the feasibility problem in all youth outcome management, and extend the privilege of client-based outcome feedback to children—to give youth a voice in the services they receive—the Child Outcome Rating Scale (CORS) (Duncan, Miller, & Sparks, 2003) was developed. This article reports the psychometric properties of the CORS for children 6-12 and the Outcome Rating Scale (ORS) (Miller & Duncan, 2000), previously validated for adults (Miller et al., 2003), for youth 13 and above for use in tracking the effectiveness of mental health and substance abuse services. The current study also sought to confirm the validity of using the youth’s voice in the delivery of services. With more and more children participating in therapy, it becomes even more important that their voices are solicited if their engagement and ultimate treatment outcome is valued.

Methods

Development of the Child Outcome Rating Scale (CORS)

Over the last several years, the Institute for the Study of Therapeutic Change (ISTC)

has developed a system for both monitoring and improving the effectiveness of treatment (Duncan et al., 2004; Duncan et al., 2003; Duncan & Sparks, 2007; Miller et al., 2003; Miller et al., 2006). The approach builds on two key findings regarding the prediction of outcome as well as the movement toward consumer involvement (Duncan, Miller, & Wampold, in press). First is the robust predictive relationship between the alliance and outcome (Norcross, in press). And second is the finding that the client's subjective experience of improvement is a reliable predictor of eventual treatment outcome (Brown et al., 2001; Howard, Kopte, Krause, & Orlinsky, 1986; Howard, Moras, Brill, Martinovich, & Lutz, 1996).

Measures of client progress and experience of the alliance can be used to “determine the appropriateness of the current treatment...the need for further treatment...[and] prompt a clinical consultation for patients who [were] not progressing at expected rates” (Howard et al. 1996, p. 1063). Providing clinicians with client-based outcome feedback dramatically increases both the effectiveness and efficiency of clinical services (Bohanske, in press; Duncan et al., in press; Lambert, in press; Miller et al., 2006). For example, Whipple et al. (2003) found that clients at risk for a negative or null outcome were less likely to deteriorate, more likely to stay longer, and *twice as likely* to achieve a clinically significant change when their therapists had access to outcome and alliance information. The efficiency of services—length of stay, productivity, cancellations, no shows, and dropouts—are also significantly improved by systematic outcome feedback (Bohanske, in press).

The ISTC outcome management system was specifically designed for utilization in partnership *with* clients, actively incorporating the now over 1100 studies that confirm the importance of client engagement and participation for treatment success (Duncan et al., in press). Such a partnership between the provider and consumer fits the ever more powerful worldwide movement to view mental health and substance abuse services from the perspective of the client—to insure that consumers have both “choice and voice” in the services they receive (Bohanske, in press). Outcome management, rather than a provider evaluation of the client, becomes a pivotal part of delivering consumer-driven services, the therapeutic relationship, and change itself.

In addition to establishing a system that is valid, reliable, and consumer driven, a major goal of the ISTC has been making the collection and use of outcome data user-friendly for both providers and consumers. As is news to no one on the front lines, and especially in the public sector, the number of forms and oversight procedures has exploded. Few have the time to devote to the repeated administration, scoring, and interpretation of lengthy measures. Brown, Dreis, and Nace (1999), for example, found that practitioners did not consider any measure that took more than five minutes to complete, score, and interpret practical. After experimenting with a number of outcome measures, we found that similar tolerance levels apply to consumers. Clients quickly tire of measures that lack obvious face validity, require more than a few minutes to complete, or appear to take away from time spent with the counselor. Low compliance rates are the most frequent result.

Feasibility, therefore, is a critical issue in outcome management. Though it may be distressing to researchers, the ease with which an instrument can be explained, completed, interpreted, and then integrated into ongoing care is much more likely to influence utilization than either validity or reliability. Seeking to make the measurement process feasible to the consumer and front-line-clinician so that outcome could be tracked and the benefits of improved effectiveness and efficiency could be realized, the ORS (Miller & Duncan, 2000) was developed as a brief alternative to the Outcome Questionnaire 45.2 (OQ) (Lambert et al., 1996). The specific items on the ORS were adapted from the three areas of client functioning assessed by the OQ; specifically, individual, relational, and social domains. Changes in these three areas are widely considered to be valid indicators of successful treatment outcome (Lambert & Hill, 2004). With regard to the specific items on the ORS, the three areas of client functioning were simply translated into a visual analog format

with instructions to place a mark on the corresponding 10 cm line, with low estimates to the left and high to the right (see Appendix 1). Research has demonstrated the reliability and validity of brief visual analog scales (e.g., Ger, Ho, Sun, Wang, & Cleeland, 1999; Zalon, 1999). In addition to their brevity, ease of administration, and scoring, such scales frequently enjoy face validity with clients typically missing from longer and more technical measures that seem distant from the client's experience.

Miller et al. (2003) reported results of an initial investigation of the reliability and validity of the ORS. Pearson product moment correlation between the ORS and the OQ yielded a concurrent validity coefficient of .58, a figure considered adequate given the brevity of the ORS. Reliability of the measure, as assessed by Cronbach's coefficient alpha, was .93. An independent study of a normative population found similar results (Bringhurst, Watson, Miller, & Duncan, 2006). Perhaps more importantly, Miller et al. (2003) also compared the feasibility of the OQ and the ORS at two community mental health centers and reported significant differences. Utilization of the 4-item ORS reached 86% at the end of one year while the 45-item OQ dropped significantly by six months and finished the year at 25%. The problem of feasibility also plagues youth outcome assessment—although reliable and valid, available outcome measures for youth are long and present an arduous task for both clinicians and clients. For example, the Ohio Scales (Ogles, Melendez, Davis, & Lunnen, 2001) "short" form is 48 items while the short form of the Youth Outcome Questionnaire is 30 items.

Seeking to simplify the assessment of youth outcome for everyday use, the CORS was developed to track the effectiveness of clinical services offered to children, as reported by children and their parents or caretakers, from age 6 through 12. The CORS (see Appendix 2) is similar in format to the ORS but contains child friendly language as well as smiley and frowny faces to aid the child's understanding. The CORS rates at a third grade reading level, making it not suitable with some children. The ORS rates at an eighth-grade reading level, making it feasible for most adolescents who seek or are referred for services. Parents and caretakers use the ORS to rate youth over 12 and the CORS for children 12 and under. Children, adolescents, parents, and caretakers have little difficulty connecting their day-to-day lived experience to the ORS/CORS and translating it into the specifics of their circumstances. Indeed, a recent qualitative study of children's experiences of the measures found a surprising level of understanding of both ORS/CORS and the therapeutic process (Crystal, 2007). The youth in this study found the measures to be practical, simple and helpful. They believed the measures opened important discussions, gave credence to their perspective, and encouraged active participation.

The Comparison Measure

The Youth Outcome Questionnaire 30 (YOQ). To assess concurrent validity of the CORS/ORS, the YOQ, the 30 item form of the YOQ-2.0, was selected because of its strong psychometric qualities and widespread use (Burlingame, Mosier, et al., 2001). It contains thirty questions that are scored on a five-point Likert scale. The YOQ provides a total score or global index of behavioral and emotional distress in a child/adolescent's life. Critical items alert clinicians to potential high-risk behaviors (e.g. suicide, substance abuse). It has a broad normative sample and has been used extensively in research. The YOQ is available in a parent-rating version which is appropriate for ages 4 through 17, and a self-report version, which is appropriate for ages 12 through 17. Most of the variance on the YOQ, like most outcome measures, can be accounted for on a single dimension: distress (Mueller, Lambert, & Burlingame, 1998). Parents or others with extensive interaction with the client complete the YOQ at intake to establish a severity baseline and then complete it at regular intervals to track the progress of the child's treatment (Burlingame, Jasper, et al., 2001).

Participants

Participants in this study were recruited from three non-clinical and three clinical sites.

Non-clinical Group

The first non-clinical site was a prevention project in a public school. Fifty-nine dyads of caretakers and youth completed the questionnaires on two separate occasions for a total of 236 administrations of the measures. The second non-clinical sample was recruited from graduate courses at Ottawa University (116 dyads of caretakers and youth). Youth and caretakers completed the measures on two separate occasions for a total of 464 administrations. The third non-clinical sample was recruited from employees of Southwest Behavioral Health (23 dyads), totaling 92 administrations of the measures over two separate occasions. Of the total 199 youth/caretaker dyads, 154 were children 12 and under and 45 were adolescents 13 and above. The under 13 group consisted of an equal amount of males and females and the 13 and above group was composed of 23 males and 22 females. Parents completed the ORS for youth 13 and above and the CORS for children 12 and under as well as the YOQ.

Clinical Group

Adolescent. The first clinical sample consisted of 1,495 adolescents treated for at least two sessions at outpatient treatment centers. Fifty-five percent of this sample was male and forty-five percent was female. A total of 9,917 youth completed ORSs were available for analysis.¹ This sample consists of adolescent ratings only.

Caretaker. The second clinical sample consisted of 1961 children and adolescents receiving publicly funded outpatient treatment services for at least two sessions within a five county rural area in Oregon. A total of 11,737 completed measures were available for analysis. This sample consists of caretaker ratings only of both children and adolescents.²

Child, Adolescent, and Caretaker. The final clinical sample consisted of 155 children and adolescents and a matched pair of 155 caretakers at a South Florida community mental health center³ that completed at least two sessions. A total of 1860 completed C/ORSs were originally available for analysis but only the first and last sessions were entered into the data base. This sample includes child, adolescent, and caretaker ratings.

Procedure

Nonclinical Group

In each non-clinical sample, participants received two concurrent administrations of the CORS (if age 6-12) or the ORS (if 13-17) and the YOQ (if over 12) over a period ranging from 10 days to 3 weeks. Caretakers were administered either the CORS or ORS, depending on the age of their child, and the YOQ.

Clinical Groups

In all clinical samples, therapists or other staff collected data as part of standard agency policy. Data from the first sample occurred over a four year span; the second over a three span; and the third represented two years of data collection. Cases where either an initial or final ORS score was missing were excluded from the sample, as were those cases that marked the maximum score throughout service.

Results

Normative Data

Table 1 displays the means and standard deviations for the non-clinical and clinical samples. As expected, a two-tailed *t*-test comparing initial ORS and CORS scores for the non-clinical and clinical samples was highly significant ($p < .0001$).

Table 1.
Comparison Of Clinical And Nonclinical Samples

<i>Sample</i>	<i>N</i>	<i>Instrument</i>	<i>Mean</i>	<i>SD</i>	
Non-clinical 199 dyads or 796 administrations		<i>(first administration mean)</i>			
	45	ORS Adolescent (28.5)	29.5	7.9	
	45	ORS Adol. Caretaker (31.0)	30.7	7.7	
	154	CORS Child (33.0)	33.4	7.0	
	154	CORS Caretaker (33.5)	33.7	6.9	
	45	YOQ Adolescent (24.6)	25.4	16.3	
	45	YOQ Adol, Caretaker (21.7)	20.9	18.7	
	154	YOQ Child Caretaker (16.0)	15.6	13.0	
	199	Total C/ORS Caretaker (32.6)	32.9	7.2	
	199	Total YOQ Caretaker (17.1)	16.8	14.7	
Clinical (over 20,000 administrations)		<i>(clinical cutoff)</i>			
	Adolescents only	1495	ORS Adolescent (27.7)	25.9	8.1
	Caretakers only	1961	C/ORS Caretaker (27.2)	21.1	7.8
	Children, Adolescents, And Caretakers	38	ORS Adolescent (27.6)	25.3	9.3
		38	ORS Adol. Caretaker (27.5)	23.9	8.9
		119	CORS Child (31.8)	30.2	7.8
		119	CORS Caretaker (29.7)	24.0	9.8
157	Total C/ORS Caretaker (29.0)	23.9	9.6		

***t*-test of CORS and ORS scores for clinical and non-clinical samples yielded significant difference ($p < .001$)**

For the non-clinical group, the mean YOQ score was comparable for both caretakers (16.8 v. 17.3) and adolescents (25.4 v. 23.4) to that reported in the test manual for the large community normative sample of the measure (Burlingame, Jasper, et al., 2001). This lends support to the premise that the non-clinical sample in this study, though relatively small, was similar. Gender differences in scores in the non-clinical sample proved to be nonsignificant. For this reason, gender specific norms are not broken out here. However, there was a tendency toward less distress reported for both females and participants who identified themselves as Hispanic (self-report and caretakers). Future research will continue to explore the possibility of gender and ethnicity-specific norms.

Reliability of the ORS/CORS

Reliability, based on 1495 adolescents and 1961 children (over 20,000 administrations of the CORS and ORS) was estimated using Cronbach's coefficient alpha, a measure of

the internal consistency of the measure. The ORS and the CORS display strong evidence of reliability, with coefficient alpha estimates of .93 and .84 respectively. These are very high coefficients of reliability for such brief measures, comparing very favorably with that reported for the YOQ, suggesting the CORS and ORS tap the factor that most if not all outcome measures tap, global distress.

Table 2
Test-Retest Correlations of The CORS, ORS, and YOQ

Measure	CORS	CTCORS	ORS	CTORS	YOQ	CTYOQ (child)	CTYOQ (adol.)
Admin.2	.60	.51	.78	.72	.75	.77	.87

An estimate of test-retest reliability was obtained by correlating the test scores at the first administration with those at the subsequent administration in the non-clinical sample. Table 2 presents the test-retest correlations for the C/ORS and YOQ. As depicted in Table 1, the mean scores changed little from administration to administration. As might be expected from a brief measure, the rest-retest reliability was generally lower for the CORS than the YOQ although the ORS compared favorably. However, test-rest reliability is an inappropriate measure of reliability for a questionnaire designed to be sensitive to client's perception of subjective change. All outcomes measures will tend show a pattern of declining correlations the first administration and each subsequent administration. This is referred to as autocorrelation. For example, Table 3 presents the autocorrelation matrices for the first five administrations of two of the clinical samples.

Table 3
Correlations Between First and Subsequent Administrations of Adolescent and Caretaker Samples

	Session 2	Session 3	Session 4	Session 5
Adolescents	0.6	0.5	0.44	0.38
<i>n=</i>	(1495)	(1420)	(1210)	(889)
Caretakers	0.63	0.57	0.51	0.48
<i>n=</i>	(1961)	(1550)	(1236)	(1009)

Validity of the ORS

Tables 4 and 5 present the concurrent validity correlation matrix for children and adolescents from the non-clinical sample that received two concurrent administrations of the ORS/CORS and the YOQ for both youth and caretakers. Note that correlations between the C/ORS and YOQ will be negative since the low score reflect low distress on the YOQ while on the ORS high scores are low distress.

Table 4
CORS and YOQ Correlation Matrix (12 and under)

	CORS	CT-CORS	CT-YOQ
CORS	1	0.63*	-0.43*
CT-CORS	0.63*	1	-0.61*
CT-YOQ	-0.43*	-0.61*	1

* represents significance

Table 5
ORS and YOQ Correlation Matrix (age 13 to 17)

	ORS	CT-ORS	YOQ	CT-YOQ
ORS	1.0	0.45*	-0.53*	-0.31
CT-ORS	0.45*	1.0	-0.46*	-0.61*
YOQ	-0.53*	-0.46*	1.0	0.69*
CT-YOQ	-0.31	-0.61*	0.69*	1.0

* represents significance

An inspection of Tables 4 and 5 reveal that the CORS and ORS were significantly related to the YOQ in all cells of the matrix demonstrating moderate concurrent validity with the well researched but much longer YOQ. Similar to the results obtained comparing the ORS for adults to the OQ (.58), a Pearson product moment correlation yielded a coefficient between the CORS/ORS and YOQ caretaker scores of .61; the ORS and YOQ completed by adolescents resulted in a .53 correlation. These correlations provide evidence of the concurrent validity of the CORS/ORS as brief alternatives for assessing global subjective distress similar to that measured by the full-scale score on the YOQ. Interestingly, the correlation was also significant between both children (.63) and adolescents (.45), and caretaker ratings, suggesting that giving voice to youth via outcome measures is supported by the evidence. Although adolescents have long had this opportunity, the CORS is the first outcome measure that taps into the perspective of children ages 6-12. Similarly, change scores (the difference between first and last session scores) of youth and their caretakers in the clinical sample were also significantly correlated.

The construct validity of the CORS/ORS rests in part on the finding that the items load on a common factor shared with other similar outcome measures. Correlations between the ORS/ORS and Caretaker CORS/ORS scores and the well-validated YOQ provide further evidence of construct validity. This aspect of construct validity rests on the assumption of an underlying unobservable trait or state that the questionnaire purports to measure. In the case of outcome questionnaires, the underlying state appears to be subjective global distress. More specific constructs such as “depression,” “anxiety,” and “interpersonal problems” can be shown to share a large percentage of variance with the global distress factor. From a purely statistical point of view, constructs based on diagnostic nomenclature such as “anxiety disorder symptoms” or “symptoms of depression” appear to have very little predictive value, as both anxiety and depression symptoms load heavily on the global distress factor.

Other indications of construct validity include: the ability of the measures to differentiate

between normative and clinical samples, and the demonstration of stability in non-clinical populations vs. change sensitivity, beyond regression to the mean, in clinical populations.

Ability to Discriminate Between Client and Nonclient Samples. Comparing first administration scores for the clinical and non-clinical groups can be used to provide evidence of construct validity. Were the CORS/ORS able to accurately discriminate between the two samples, initial scores would be expected to be significantly lower (more distressed) for the clinical group. The results presented in Table 1 confirm that this is the case.

Sensitivity to Change. If valid, the ORS/CORS should reflect change following psychotherapy, but remain stable in an untreated population. Therefore, it was expected that ORS/CORS scores in the clinical sample would increase while those in the non-clinical sample would vary only minimally from a pre- and post-test. Such a finding would provide additional evidence of construct validity for the instrument.

Table 6
Change Sensitivity in Clinical Samples

<i>Clinical Sample</i>	<i>First Session Mean (SD)</i>	<i>Last Session Mean (SD)</i>	<i>Pre-post Change (Significance)</i>
ORS Adolescent N=1495	25.9 (8.1)	33.6 (6.5)	7.9 (p< .001)
C/ORS Caretaker (CT) N=1961	21.1 (7.8)	24.4 (7.7)	3.3 (p< .001)
ORS Adolescent CT Adolescent N=38	25.3 (9.3) 23.9 (8.8)	29.7 (10.2) 29.0 (8.1)	4.4 (p< .05) 5.1 (p< .001)
CORS (12 and under) CT (12 and under) N=119	30.2 (7.8) 24.0 (9.8)	35.2 (6.3) 31.2 (7.6)	5.0 (p< .001) 7.2 (p< .001)

An inspection of Table 6 reveals that across all three samples with children, adolescents, and caretakers, this was indeed the case. A *t*-test for correlated samples tested the hypothesis that CORS/ORS scores would increase following therapy intervention. As expected, the *t*-test between the mean of the client pre-test scores and their post-test scores revealed statistically significant improvement. Both self-rated and caretaker-rated measures showed significant improvement between pre- and post-test scores. (p<.01; one-tailed *t*-test) The caretaker-rated measures averaged 6.7 points improvement, compared to 4.8 on the self-rated measures. The mean change on the self rated measures was likely limited due to the skewed nature of the scores at pre-test.

Conversely, a *t*-test between mean pre- and post- CORS/ORS test scores from the non-clinical sample proved nonsignificant. Therefore, the CORS/ORS was sensitive to change in those clients receiving psychotherapy and relatively stable for those not receiving intervention.

Sensitive to Change beyond Regression to the Mean. To test if the observed changes exceeded regression to the mean a time-reversed regression was employed (Cohen & Cohen, 1983). Pure regression to the mean is a measurement artifact that is time symmetrical, meaning that the equation used to predict the last score from the first score is the same if time is reversed and the last score is used to predict the first score. If the change between the first and last score is greater in the forward prediction than in the time-reversed prediction, then the observed change exceeded regression to the mean.

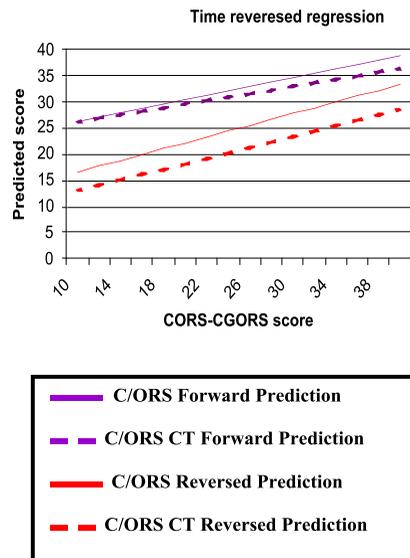
Figure 1. Regression to the Mean

Figure 1 presents the results of the analyses. The x-axis displays C/ORS child/adolescent and caretaker scores. For the purpose of this analysis children and adolescents were considered together. The y-axis displays the predicted score resulting from the regression analysis. The purple lines (solid and dotted) represent the forward predictions for each measure, while the red lines are the time-reversed predictions. The results clearly demonstrate that change on each measure exceeded regression to the mean.⁴ Strikingly, the forward predictions are virtually identical for the CORS/ORS and caretaker ratings. Even though the distributions of these scores on these measures differed in this sample, with youth reporting less distress than their parents, the underlying relationship between intake score and final score (or change score) is very similar, and another indication that the voice of youth should be incorporated.

Discussion

All the problems typically associated with brief self-report tests (Boulet & Boss, 1991) apply to the CORS/ORS; for example, interpretation relies on clients' accurate assessment of their levels of distress and there are no controls for response sets like social desirability. Additionally, the CORS/ORS does not measure nor is intended to identify clinical risk factors such as suicide or alcohol or drug use. Research with more diverse clinical and non-clinical samples is underway and should further identify the strengths and weakness of the measure. Clearly, evaluation of outcome via the ORS and CORS is far from comprehensive and does not contain multiple perspectives (e.g., therapists, outside judges, community criteria, etc.).

Although a short measure like the CORS/ORS cannot be expected to achieve the same precision or depth of information as a longer measure like the YOQ, this study found that the CORS/ORS has moderate validity and solid reliability. The CORS for children under 13 and their caretakers and the ORS for adolescents and their caretakers provide a brief measure of global distress suitable for assessing treatment outcomes. The measure shows unusually high reliability for a 4-item questionnaire, comparing favorably to well establish outcome measures containing many more items. The high coefficient alpha provides evidence of construct validity; the four items all correlate with one another and appear to measure the broad construct of global distress.

The non-clinical sample was used to assess the concurrent validity of the CORS/ORIS. The overall correlation with the YOQ demonstrated that the C/ORIS is moderately related to this gold standard of self-report scales. Lower scores (more distressed) at the first administration were anticipated for the clinical sample as compared to the non-clinical. The difference found between the clinical and non-clinical samples suggests that the CORS/ORIS measures what it purports to: psychological distress. Changes in client scores between pre-and post-test as compared to the stable scores for the non-clinical sample also provide evidence of construct validity for the C/ORIS

This study also addressed the issue of including the youth's voice in clinical decision making—whether a child/adolescent self-report measure is an appropriate source of information. Developmental considerations, such as verbal skills and reading ability, motivation for treatment, social desirability, and perception of the problem (i.e. whether the child sees a problem v the caretaker perspective) have all combined to make caretaker report the norm for children under twelve and the probability for adolescents. The assumption is that parents will be the most reliable source for obtaining data about a child's functioning. In adolescents, Achenbach and Edelbroch (1991) suggest that the major differences between parent and self-report are: 1) adolescents under-report problems as compared to parents and 2) parents are better sources regarding objective behaviors (e.g. oppositional attitude, school failures, etc) while adolescents are more accurate about their subjective states (moods, feelings, etc). It is our view that any a priori dismissal of the child (too young) or adolescent's (objective behaviors) perspective is seriously misguided. Children, much like adults diagnosed with a severe mental illness, have had a profound absence of voice in the delivery of services under the justification that they do not know what is best for them. This injustice is compounded by the fact that youth are most often mandated for services, and thus are subjected to the whims, well-intended as they are, of the adults who decide on their behalf. This study demonstrated that while youth do rate themselves higher (less distressed) than their caretakers, their views are nonetheless positively correlated with caretakers and are reliable and valid markers of treatment of success. Youth scores increased with psychotherapy just as the caretakers' scores.

The systematic incorporation of the youth's voice may be criticized as naïve or ill-informed. However, treating young people as reliable informants and helping them make sense of their experience in ways that generate hope and engagement is well grounded in empirical evidence. Recent meta-analyses of the child outcome literature indicate that no one approach is superior to another for resolving child problems (Miller, Wampold, & Varhely, in press). We simply do not know what will be helpful for the individual client unless we obtain feedback about the effectiveness of any intervention from the client's point of view. Additionally, research confirms the pivotal role the alliance plays in the outcome of child intervention (Shirk & Karver, 2003). When children and adolescent experience that their opinions matter, participation is enhanced, enabling the practitioner to harness the most powerful therapist provided-means to promote change—the alliance.

The brevity of the measures greatly aids the partnership process—feasibility is critical to outcome monitoring and management. The length of the YOQ and its lack of face validity made the completion of this study doubtful at times. The original study included 7 normative sites, but they quickly dropped out as they discovered the difficulty in getting children, adolescents, and their families to complete more than one pass at the YOQ. A few of the sites offered incentives: raffles, dinners, and payment, but the result was the same—a majority of those who volunteered dropped out. In one school site, following a donation to the school, 500 youth/parents dyads volunteered for the study. At the first assessment, only 200 completed the measures. Of that 200, only 25 returned for a second assessment. In total, over 2500 research packets were disseminated which finally resulted in a non-

clinical sample of 199 dyads. We believe this illustrative of the feasibility of the measures involved.

On the practitioner side of things, many therapists see outcome measurement as an “add-on” separate from actual clinical work and relevant only to management and other overseers. In addition to wanting measures to be brief, easy to integrate, and have face validity, counselors want measures that are clinically useful. Is the measure intended to improve the effectiveness of rendered services or merely monitor them? Most if not all other youth outcome measures were developed primarily as pre-post and/or periodic outcome measures. Such instruments, like the 48 item Ohio Scales, provide an excellent way to measure program effectiveness but are not feasible to administer frequently, and therefore, do not provide real-time feedback for immediate treatment modification before clients drop out or suffer a negative outcome—in short they are not clinical tools as much as they are oversight tools.

Longer measurement systems largely intended for oversight can also create significant management problems. In reaction to a managed care company’s introduction of the Outcome Questionnaire 30, just 30 items, it was recently reported in the *New England Psychologist* (Hanlon, 2005) that providers complained about its length and frequent administration, that it cut into sessions and increased workload, and that some items were intrusive. The response by clinicians was so severe that it led the State Psychological Association president to say, “*I have never seen such negative reaction from providers*” (Hanlon, 2005, p. 11). This is not an infrequent reaction in our experience.

The ORS/CORS was designed as a clinical *and* outcome tool to provide immediate feedback to both clients and providers to improve the effectiveness of services, and as a way to measure outcome at individual, program, and agency levels. Given their feasibility, the ORS/CORS can provide immediate feedback not only based on client scores but also in comparison to normative trajectories of change of a large and growing clinical data base (over 300,000 administrations). In addition, the measures were designed to be used in collaboration with clients, to encourage a partnership between the client and therapist for monitoring the effectiveness of services. Accountability becomes a joint endeavor, integral to alliance building, rather than simply more paperwork.

To be sure, the ORS and CORS are weaker psychometrically than the YOQ and other longer measures. Neither do these brief visual analogue scales offer the same breadth of assessment as the longer scales (e.g., the absence of “critical items” related to suicide or alcohol and drug use). At the same time, a measure that goes unused is worthless regardless of its strengths. In the real world of clinical practice, measuring outcome means striking a balance between the competing demands of validity, reliability, and feasibility. The development of the ORS, and subsequently the CORS reflect the ISTC’s attempt to find such a balance.

Moreover, the ORS and CORS represent our efforts to bring consumers into the loop in decisions about their care to allow a true partnership between clients and providers of service—to improve outcome one client at a time by assigning those we serve key roles. The current study demonstrated that young people can also be given a voice in mental health and substance abuse services, and assume their rightful roles in the forefront of their own change.

References

- Achenbach, T. M., & Edelbroch, C. (1991). *The child behavior checklist and revised child behavior profile*. Burlington, VT: University Associates in Psychiatry.
- Bohanske, R.T. (in press). Transforming public behavioral health care. In B. Duncan, S.

- Miller, & B. Wampold (Eds.), *The heart and soul of change: Delivering what works*, 2nd ed. Washington, DC: APA Press.
- Boulet, J., & Boss, M. (1991). Reliability and validity of the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology*, 3(3), 433–437.
- Bringham, D.L., Watson, C.W., Miller, S.D., & Duncan, B.L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, 5, 23-30.
- Brown, J., Dreis, S., & Nace, D. (1999). What really makes a difference in psychotherapy outcome? Why does managed care want to know? In M. Hubble, B. Duncan, & S. Miller (Ed.), *The heart and soul of change* (pp. 389–406). Washington, DC: APA Press.
- Brown, G. S., Burlingame, G. M., Lambert, M. J., Jones, E., & Vaccaro, J. (2001). Pushing the quality envelope: A new outcomes management system. *Psychiatric Services*, 52 (7), 925-934.
- Burlingame, G. M., Jasper, B. W., Peterson, G., Wells, M. G., Lambert, M. J., Reisinger, C. W. & Brown, G. S. (2001). *Administration and Scoring Manual for the Y-LSQ*. Stevenson, MD: American Professional Credentialing Services LLC and Pacificare Behavioral Health, Inc.
- Burlingame, G. M., Mosier, J. I., Wells, M. G., Atkin, Q. G., Lambert, M. J., Whoolery, M., & Latkowski, M. (2001). Tracking the influence of mental health treatment: The development of the Youth Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 8, 361-379.
- Cohen, J., & Cohen, T. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crystal, S.C. (2007). Exploring children's experiences of the use of outcome measures in therapy. Unpublished masters dissertation, University of Birmingham, Birmingham, UK.
- Day, C., Carey, M., & Surgenor, T. (2006). Children's key concerns: Piloting a qualitative approach to understanding their experience of mental health care. *Journal of Clinical Child Psychology and Psychiatry*, 11 (1), 139-155.
- Duncan, B. L., Miller, S.D., Sparks, J. A. (2003). *The Child Outcome Rating Scale*. Authors: Ft. Lauderdale, FL.
- Duncan, B. L., Miller, S.D., Sparks, J.A. (2004). *The heroic client: A revolutionary way to improve effectiveness*, rev. ed.. San Francisco: Jossey Bass.
- Duncan, B.L., Miller, S.D., & Wampold, B.E. (Eds.) (in press). *The heart and soul of change: Delivering what works*, 2nd ed.. Washington, DC: APA Press.
- Duncan, B.L., & Sparks, J.A. (2007). *Heroic clients, heroic agencies: Partners for change*, rev. ed. www.talkingcure.com: ISTC Press
- Ger, L. P., Ho, S. T., Sun, W. Z., Wang, M. S., & Cleeland, C. S. (1999). Validation of the Brief Pain Inventory in a Taiwanese population. *Journal of Pain & Symptom Management*, 18(5), 316–322.
- Hanlon, P. (2005). PacifiCare screening tool, policies raise concerns. *New England Psychologist*, 13, 1,11-12.
- Howard, K.I., Kopte, S.M., Krause, M.S., & Orlinsky, D.E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, 41, 159-164.
- Howard, K., Moras, K., Brill, P., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, 51(10), 1059–1064.
- Kazdin, A. E. (2004). Psychotherapy for children and adolescents. In M. J. Lambert (Ed.). *Bergin and Garfield's handbook of psychotherapy and behavior change*, 5th ed. (pp. 543-589). New York: Wiley.

- Lambert, M.J. (in press). Yes, it is time for clinicians to routinely track outcome. In B. Duncan, S. Miller, & B. Wampold (Eds.), *The heart and soul of change: Delivering what works, 2nd ed.* Washington, DC: APA Press.
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3 (4), 249-258.
- Lambert, M. J., Okiishi, J. C., Finch, A. E., & Johnson, L. D. (1998). Outcome assessment from conceptualization to implementation. *Professional Psychology: Research and Practice*. Vol. 29, No. 1, 63-70.
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians routinely to track patient outcome? A meta-analysis. *Clinical Psychology*, 10, 228-301.
- Lambert, M. J., & Hills, C. E. (2004). Methodological issues in assessing psychotherapy processes and outcomes and processes. In M. J. Lamberts (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior and change, 5th ed.* (pp. 84-136). New York: John Wiley.
- Lambert, M. J., Whipple, J., Smart, D., Vermeersch, D., Nielsen, S., & Hawkins, E. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11(1), 49-68.
- Miller, S. D., & Duncan, B. L. (2000). *The Outcome Rating Scale*. Chicago: Author.
- Miller, S. D., Duncan, B.L., Brown, J., Sorrell, R., & Chalk, M. B. (2006). Using outcome to inform and improve treatment outcomes. *Journal of Brief Therapy*, 5, 5-22.
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief, visual, analog measure. *Journal of Brief Therapy*, 2, 91-100.
- Miller S., Duncan, B., Sorrell, R., & Brown, J. (2005). The Partners for Change Outcome Management System. *Journal of Clinical Psychology: In Session*, 61, 199-208.
- Miller, S.D., Wampold, B.E., & Varhely, K. (in press). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*.
- Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, 70(2), 248-262.
- Norcross, J.C. (in press). The therapeutic relationship. In B. Duncan, S. Miller, & B. Wampold (Eds.), *The heart and soul of change: Delivering what works, 2nd ed.* Washington, DC: APA Press.
- Ogles B.M., Melendez G., Davis D.C., & Lunnen K.M (2001). *Journal of Child and Family Studies*, 10 (2), 199-212.
- Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 71(3), 452-464.
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology*, 50(1), 59-68.
- Zalon, M. L. (1999). Comparison of pain measures in surgical patients. *Journal of Nursing Measurement*, 7(2), 135-152.

Appendix 1: Outcome Rating Scale (ORS)

Name _____ Age (Yrs): _____ Sex: M / F
 Session # _____ Date: _____
 Who is filling out this form? Please check one: Self _____ Other _____
 If other, what is your relationship to this person? _____

Looking back over the last week, including today, help us understand how you have been feeling by rating how well you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels. *If you are filling out this form for another person, please fill out according to how you think he or she is doing.*

Individually (Personal well-being)

|-----|

Interpersonally (Family, close relationships)

|-----|

Socially (Work, school, friendships)

|-----|

Overall (General sense of well-being)

|-----|

Institute for the Study of Therapeutic Change

www.talkingcure.com

© 2000, Scott D. Miller and Barry L. Duncan

Note: Working copies of the measures are available for free download from www.talkingcure.com.

Appendix 2: Child Outcome Rating Scale (CORS)

Name _____ Age (Yrs): _____

Sex: M / F _____

Session # _____ Date: _____

Who is filling out this form? Please check one: Child _____ Caretaker _____

If caretaker, what is your relationship to this child? _____

How are you doing? How are things going in your life? Please make a mark on the scale to let us know. The closer to the smiley face, the better things are. The closer to the frowny face, things are not so good. *If you are a caretaker filling out this form, please fill out according to how you think the child is doing.*

Me

(How am I doing?)



I-----I



Family

(How are things in my family?)



I-----I



School

(How am I doing at school?)



I-----I



Everything

(How is everything going?)



I-----I



Institute for the Study of Therapeutic Change

www.talkingcure.com

(Endnotes)

1. The data from this site was released on the condition of total anonymity.
2. This sample from Accountable Behavioral Health Alliance used a derivative of the ORS/CORS called the Oregon Change Index (OCI). The ORS/CORS utilizes a 10-centimeter visual analog scale with anchors at the extremes of the scale. The OCI is scored on a 10-point Likert scale, also with anchors at the extremes. A Likert scale with discrete points was used to reduce labor costs associated with scoring and data entry. Both methods of presenting and scoring the scales appear to create very similar results, though differences in response patterns to a visual analog versus 10-point Likert scale may result in small differences in the psychometric properties of the scores produced.
3. The Center for Family Services of Palm Beach County, Inc. is a not-for-profit family services agency serving Palm Beach County of South Florida. The agency provides an array of services including individual and family counseling, substance abuse treatment, sexual abuse and domestic violence treatment, EAP services, homeless assistance/shelter, a school readiness program, and home-based family therapy to families at risk.
4. Cautionary note: The skewed nature of the C/ORS scores violates the assumption of normality underlying the use of regression techniques. While these techniques are generally robust and not overly sensitive to deviations from normality, some caution should be used when interpreting regression results for the C/ORS data. A sample with less skewed scores may yield different results.

