

# **The Outcome Rating Scale: A Preliminary Study of the Reliability, Validity, and Feasibility of a Brief Visual Analog Measure**

**Scott D. Miller, PhD**

**Barry L. Duncan, PsyD**

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

**Jeb Brown, PhD**

*Center for Clinical Informatics  
Salt Lake City, Utah*

**Jacqueline A. Sparks, PhD**

**David A. Claud, MS**

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

Industry-wide, there is a trend toward making outcome evaluation a routine part of therapeutic services. Although various multidimensional assessments of outcome are valid and reliable, their methodological complexity, length of administration, and cost often render them infeasible for many service providers and settings. The present article describes the development and validation of an ultra-brief outcome measure, the Outcome Rating Scale (ORS). The instrument's psychometric properties are examined and reported for both clinical and nonclinical samples. Based on experience with the instrument at the various sites in the study, the feasibility of the scale is considered. Results indicate that the ORS represents a balanced trade-off between the reliability and validity of the longer measures, and the feasibility of this brief scale. Results and implications for clinical practice and future research are discussed.

**T**he systematic evaluation of outcome is becoming increasingly more routine in the provision of mental health and therapeutic services. Policy makers, third-party payers, government agencies, and consumers are concerned that limited health care dollars be spent on treatments that work. While therapists may feel singled out in this process, observers

note that interest in outcome is not specific to mental health, but rather is part of a worldwide trend (Andrews, 1995; Humphreys, 1996; Lambert, Okiishi, Finch, & Johnson, 1998; Sanderson, Riley, & Eshun, 1997). As researchers Brown, Dreis, and Nace (1999) point out:

*In the emerging environment, the outcome of the service rather than the service itself is the product that providers and payers have to market and sell. Those unable to systematically evaluate the outcome of treatment will have nothing to sell to purchasers of healthcare services. (p. 393)*

Presently, a variety of approaches exist for evaluating the outcome of psychotherapy. Most such efforts draw on well-established measures, both clinician and client rated, as well as observer ratings, physiological indices, and environmental information (Lambert, Ogles, & Masters, 1992). While these multidimensional assessments of outcome are valid and reliable, their methodological complexity, length of administration, and cost often render them infeasible for many service providers and settings. The average clinician's caseload is already overloaded with paperwork or other nondirect service-related activities (e.g., phone calls, team meetings, treatment planning, progress notes, and so on [Duncan & Miller, 2000]). Brown and colleagues (1999), for example, found that the majority of clinicians did not consider any measure or combination of measures that took more than 5 minutes to complete, score, and interpret practical. As a result, a strong argument can be made for adopting measures that are brief in addition to being reliable and valid.

In addition to responding to funding source requirements and consumer demand, recent studies have explored how outcome evaluations can be used on an ongoing basis both to inform clinical decision making and enhance treatment effects. A growing body of research indicates, for example, that the client's subjective experience of change early in the treatment process is one of the better predictors of treatment outcome (Duncan & Miller, 2000; Howard, Moras, Martinovich, & Lutz, 1996; Lambert & Bergin, 1994). In two studies, Lambert and colleagues (2001) and Whipple and colleagues (2003) found that incorporating outcome information into therapy resulted in a 65% improvement in the success of cases most at risk for a negative or null outcome. In another study of over 3,000 cases at a single agency, ongoing use of outcome information over 1 calendar year resulted in a 150% improvement in overall effectiveness (Miller, Duncan, Brown, Sorrell, & Chalk, 2003).

The present article describes the development and validation of an ultra-brief outcome measure, the Outcome Rating Scale (ORS; Miller & Duncan, 2000; see Appendix). The instrument's psychometric properties are examined and reported for both clinical and non-clinical samples. Based on experience with the instrument at the various sites in the study, the feasibility of the scale is considered. Results and implications for clinical practice and future research are discussed.

## METHODS

### Development of the Outcome Rating Scale (ORS)

The ORS was developed as a brief alternative to the Outcome Questionnaire 45.2 (Lambert, Hansen, et al., 1996). The first two authors were familiar with the latter instrument, having used it in their own practices, as well as in research and consultation with numerous mental health agencies and a number of third-party payers. In virtually all instances, complaints regarding the length of time needed to complete the longer measure, as well as the size of the print, and content of the questions were quick to surface among clinicians and clients. In spite of the availability of computerized and telephonic versions, as well as the possibility of

intermittent rather than consistent use of the scale from session to session, problems with administration persisted. In busy clinics, for example, a single client showing up late could wreak havoc with the schedule. Use of the instrument in emerging telephonic and Internet-based services was deemed impossible for similar reasons.

The specific items on the ORS were adapted from the three areas of client functioning assessed by the OQ-45.2; specifically, individual, relational, and social. Changes in these three areas are widely considered to be valid indicators of successful treatment outcome (Kazdin, 1994; Lambert, Burlingame, et al., 1996; Lambert & Hill, 1994).

Research has demonstrated the reliability and validity of ultra-brief visual analog scales in several areas including: assessment and management of pain (see Ger, Ho, Sun, Wang, & Cleeland, 1999; Radbruch et al., 1999; Zalon, 1999); perceived quality of care (see Arneill & Devlin, 2002); psychoeducation (see Dannon, Iancu, & Grunhaus, 2002); health states preferences (Torrance, Feeny, & Furlong, 2001); and even the assessment of change in response to medical treatments (see Grunhaus, Dolberg, Polak, & Dannon, 2002). 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. With regard to the specific items on the ORS, the three areas of client functioning were simply translated into visual analog format, with instructions to place a hash mark on the corresponding 10 cm line, with low estimates to the left and high to the right.

## Participants

Participants in this study were recruited from one clinical and one non-clinical population.

**Nonclinical Group.** The nonclinical group consisted of 86 participants made up of graduate masters-level students ( $n = 78$ ), therapists and staff ( $n = 9$ ) working at a community family service agency (CFS). Members of this group were either part-time or full-time residents from South Florida communities who ranged in age from 22 to 65, and were of mixed gender, socioeconomic, and ethnic backgrounds.

**Clinical Group.** Clinical data were collected from clients at the CFS ( $n = 435$ ). CFS clients were enrolled in traditional office-based counseling and came to the agency presenting a typical range of initial difficulties. Among this sample were clients entering the program as employee assistance program (EAP) referrals. The CFS client sample was limited to adults aged 18 and above who had received a minimum of three clinical sessions and a maximum of 10 clinical sessions. This parameter was based on previous studies indicating that change in therapy is likely to occur after the completion of three sessions, with diminishing rates of change beyond 10 (Brown et al., 1999; Howard et al., 1986). Clients from the agency's substance abuse recovery program were excluded. Since substance abuse clients, in general, were mandated, outcome measurements might reflect concerns other than actual progress in counseling (e.g., referral issues) (see Lambert, Burlingame, et al., 1996).

## Measures

**The Outcome Questionnaire 45.2.** The OQ-45.2 is a 45-item self-report scale designed for repeated measurement of client functioning through the course of therapy and at termination. The measure has high internal consistency (.93) and test-retest reliability (.84). Moderate to high validity coefficients have been reported between the scale and other well-established measures of depression, anxiety, and global adjustment, for example, the SCL-90-R, Zung Depression and Anxiety Scales, Beck Depression Inventory, and Taylor Manifest Anxiety Scale (Lambert, Hansen, et al., 1996). Factor analysis of the OQ-45.2 failed to support

construct validity for the three subscales (individual, interpersonal, and social role functioning). The research indicates that most of the variance can be accounted for on a single dimension: distress (Mueller, Lambert, & Burlingame, 1998).

With increasing frequency, the OQ-45.2 is being used in psychotherapy outcome research. The instrument has proven particularly useful in documenting the effect of interventions due to therapy as it has been shown to be sensitive to change in a treated population while remaining stable in a nontreated population (Lambert, Burlingame, et al., 1996). Two studies have further documented the scale's ability to identify and improve the chances of success in cases at risk for a negative or null outcome (Lambert et al., 2001; Whipple et al., 2003).

## Procedure

**Nonclinical Group.** Participants in the nonclinical group received four concurrent administrations of the ORS and OQ-45.2 measures over a period ranging from 10 days to 5 weeks. Six graduate students in the sample were tested in a classroom setting, with a proctor administering the instruments. They were retested 1 to 3 weeks following the initial testing period for an additional three to four concurrent tests. One subject student completed only the first administration and therefore was excluded from the analysis. Therapist and staff participants completed the ORS and OQ-45.2 either at home or in their offices. The four test-retest administrations were separated by 1 day to 2 weeks.

**Clinical Group.** At the CFS, therapists or the staff receptionist collected ORS data from the clinical sample. The measure was introduced as a means of determining and improving therapy outcome. Administration of the ORS was part of standard agency policy. Only first and last (pre- and post-) ORS test scores were entered into the agency database. Pre- and post-ORS data were collected for those cases closed from February 2002 through February 2003. Cases where either an initial or final ORS score was missing were excluded from the sample. A total sample of 435 adults between the ages of 18 and 83 were available for analysis.

## RESULTS

### Normative Data

Table 1 displays the means and standard deviations for the nonclinical and clinical samples. As expected, a two-tailed *t*-test comparing initial ORS scores for the nonclinical and clinical samples was highly significant ( $p < .00001$ ).

For the nonclinical sample, the mean OQ-45.2 score was comparable to that reported in the test manual for the large community normative sample of the measure (Lambert, Hansen, et al., 1996). This lends support to the premise that the nonclinical sample in this study, though relatively small, was similar. As in the case of the normative sample reported for the OQ-45.2 (Lambert, Burlingame, et al., 1996), age and sex differences in scores in the nonclinical sample proved to be nonsignificant. For this reason, sex and age specific norms are not broken out here.

TABLE 1. COMPARISON OF CLINICAL AND NONCLINICAL SAMPLES

	<i>n</i>	Instrument	<i>M</i>	<i>SD</i>
Nonclinical	86	ORS	28.0	6.8
		OQ-45	46.6	16.2
Clinical	435	ORS	19.6	8.7

Two-tailed *t*-test comparison of ORS scores for nonclinical and clinical samples.  
 $p < .0001$ .

For the clinical sample, significant differences in intake scores were found between men and women ( $p < .001$ ). Table 2 displays the means and standard deviations of the ORS scores broken out by sex.

### Reliability of the ORS

Both test-retest and internal consistency reliability were evaluated using the nonclinical sample ( $N = 86$ ). Cronbach's coefficient alpha was calculated as the estimate of internal consistency. Coefficient alpha ranged from .87 at the first administration to .96 at the third and fourth administration. Coefficient alpha for all administrations ( $N = 336$ ) was .93.

Coefficient alpha (.93) for the ORS compared favorably with that reported for the OQ-45.2. As a rule, one would expect a measure with only four items to have a lower reliability than a measure consisting of 45 items. This high degree of internal consistency reflects the fact that the four items correlate quite highly with one another, indicating that the measure perhaps can best be thought of as a global measure of distress rather than one possessing subscales for separate dimensions.

An estimate of test-retest reliability was obtained by correlating the test scores at the first administration with those at each subsequent administration. Table 3 presents the test-retest correlations for the ORS and OQ-45.2. As would be expected from an ultra-brief measure, the test-retest reliability was significantly lower than the OQ-45.2. However, interpretation of test-retest reliability in a measure meant to be sensitive to change can be difficult. For example, the lower correlations could be due to the measure being more sensitive to small changes in the individual's global sense of well-being.

### Validity of the ORS

**Concurrent Validity.** Concurrent validity was computed using Pearson product-moment correlations (Cohen & Cohen, 1983) between the ORS total score and OQ-45.2 total score, as well as subscale scores on the nonclinical sample. Table 4 displays the correlation coefficients at each administration. In order to explore the relationship between ORS items and OQ-45.2 subscales, the data from all four administrations were combined to create a sample

TABLE 2. COMPARISON OF MALES AND FEMALES IN CLINICAL SAMPLE

	<i>n</i>	<i>M</i>	<i>SD</i>
Males	157	22.3	8.5
Females	323	18.9	8.7

Two-tailed *t*-test comparison of ORS scores male and females in clinical sample.  
 $p < .001$ .

TABLE 3. TEST-RETEST CORRELATIONS

	2nd Administration ( <i>n</i> = 86)	3rd Administration ( <i>n</i> = 86)	4th Administration ( <i>n</i> = 77)	$\alpha$
ORS	.66	.58	.49	.93
OQ-45.2	.83	.75	.74	.93

TABLE 4. ORS AND Q-45.2 COEFFICIENT OF CORRELATIONS

	1st Administration ( <i>n</i> = 86)	2nd Administration ( <i>n</i> = 86)	3rd Administration ( <i>n</i> = 86)	4th Administration ( <i>n</i> = 77)
	.69	.53	.54	.56

of 335 paired administrations for the 86 subjects. Table 5 displays the correlation coefficients between the items of the ORS and the subscales of the OQ-45.2, as well as the correlations with the total scores.

An inspection of Table 5 reveals a consistent pattern of moderately strong correlations between the ORS items and the OQ-45.2 subscales, as well as total scores. However, there is little evidence that specific items correlate significantly more with specific subscales on the OQ-45. In all cases, the individual items on the ORS correlated more highly with the OQ-45.2 total score than any of the OQ-45.2 subscales.

The overall correlation between the ORS and OQ-45.2 total scores is .59—a moderate indication of concurrent validity. Here again, the ultra-brief nature of the ORS and the novel analog scoring method may be responsible. Indeed, though modeled on the OQ, it is not reasonable to expect very high coefficients of correlation between the two measures given the shorter nature of the ORS. Nonetheless, the correlation is respectable and does provide evidence that the ORS is an ultra-brief alternative for assessing global subjective distress similar to that measured by the full-scale score on the OQ-45.2.

**Sensitivity to Change.** If valid, the ORS should reflect change following psychotherapy interventions, but remain stable in an untreated population (Lambert & Bergin, 1994). Therefore, it was expected that ORS scores in the clinical sample would increase while those in the nonclinical sample would vary only minimally from a pre- and posttest. Such a finding would provide some evidence of construct validity for the instrument.

CFS clinical sample posttest data were gathered at session 10. For purposes of comparison, change scores were calculated for the 77 individuals in the nonclinical sample that had a total of four administrations of the ORS. Change was determined by comparing the total score at first administration to that of the final administration.

A *t*-test for correlated samples tested the hypothesis that ORS scores would increase following therapy intervention. As expected, the *t*-test between the mean of the client pretest total scores and their posttest total scores revealed statistically significant improvement. Conversely, a *t*-test between mean pre- and post-ORS test scores from the nonclinical sample proved nonsignificant. Therefore, the ORS was sensitive to change in those clients receiving psychotherapy intervention and relatively stable for those not receiving intervention. These data are presented in Table 6.

TABLE 5. COEFFICIENTS OF CORRELATION BETWEEN ORS ITEMS AND OQ-45.2 SUBSCALES

	OQ-45 Subjective Distress	OQ-45.2 Interpersonal	OQ-45.2 Social Role	OQ-45.2 Total
ORS individual	.53	.50	.41	.56
ORS interpersonal	.46	.58	.35	.53
ORS social role	.42	.46	.34	.47
ORS overall	.55	.54	.42	.59
ORS total	.54	.57	.41	.59

TABLE 6. PRE-POST CHANGE FOR NONCLINICAL AND CLINICAL SAMPLES

	<i>n</i>	Pretest <i>M</i> ( <i>SD</i> )	Posttest <i>M</i> ( <i>SD</i> )	Two-Tailed <i>t</i> -Test Pre-Post Comparison
Nonclinical	77	27.9 (6.8)	29.4 (7.0)	$p > .1$
Clinical	435	19.6 (8.7)	25.7 (8.7)	$p > .00001$

**Ability to Discriminate Between Client and Nonclient Samples.** Comparing pretest scores for the clinical and nonclinical groups also can be used to provide evidence of construct validity. Were the ORS able to accurately discriminate between the two samples, initial scores would be expected to be significantly lower for the clinical group. The results presented in Table 1 confirm that this is the case.

### Feasibility of the ORS

Feasibility of an outcome instrument involves the likelihood that the instrument, in fact, will be used. In clinical settings, feasibility is the degree to which an instrument can be explained, completed, and interpreted quickly and easily. If outcome measures do not meet the time demands of actual clinical practice, they may be met with resistance by staff and clients alike. In addition, measures that are difficult to score or appear irrelevant to clients and therapists are less likely to be used.

Examining compliance rates over time in clinical sites with similar clients and mandates can be used to assess the feasibility of the ORS. Table 7 shows that at CFS, a compliance rate of 86% was achieved at the end of one year. Implementation of the OQ-45.2 in a similar outpatient setting had markedly different results. Utilization data were obtained from Family Therapy Associates (FTA) at Nova Southeastern University from June 1998 to June 1999. During that time period, a pilot study examining the effectiveness of incorporating client feedback in therapy was conducted using the OQ-45.2. FTA is a community mental health agency very similar in nature and scope to CFS. While FTA therapists were not mandated to utilize the outcome instrument, they were part of an ongoing research team and compliance was expected. Therapists were masters and doctoral students with close supervision and support provided throughout the research project. In spite of ongoing training and encouragement, however, use of the OQ dropped at 6 months and finished the year at 25%.

## DISCUSSION

This article reports on the development of an ultra-brief outcome scale. The ORS was designed for use by clinicians to assess change in clients following psychological intervention. Although a short measure cannot be expected to achieve the same precision or depth of information as a longer measure like the OQ-45.2, this study found that the ORS has adequate validity, solid reliability, and high feasibility.

Using a nonclinical sample, results demonstrated that the ORS possesses moderate stability, as reflected by the test-retest coefficients. Future research should focus on the stability of the ORS with clinical samples prior to clients undergoing therapy and over longer periods of time with normal controls. The internal consistency for the nonclinical sample was very high for the overall ORS, as well as for the subscale scores. The high intercorrelations among the subscale scores suggest a single underlying factor, not unlike the single underlying factor of distress associated with the OQ-45.2.

The nonclinical sample also was used to assess the validity of the ORS. Although not as strong as hoped, the overall correlation with the OQ demonstrated that the ORS is moderately related to this gold standard of self-report scales. Given that 45 items were reduced to four, a .59 indicator of concurrent validity meets expectations. However, a larger and more diverse nonclinical example may have yielded a higher correlation.

TABLE 7. PERCENTAGE OF COMPLIANCE WITH ORS AND OQ IN COMMUNITY AGENCIES

	1 Month	6 Months	12 Months
ORS	3%	61%	89%
OQ-45.2	33%	25%	25%

Lower ORS scores were anticipated for the clinical sample at first administration. The difference found in the study suggests that the ORS measures what it purports to: psychological distress. Changes in client scores between pre and posttest as compared to the stable scores for the nonclinical sample provide preliminary evidence of construct validity for the ORS. It is curious that females scored significantly lower than males in the clinical sample. No explanation for this difference can be offered at present. Ongoing research is currently examining the issue in greater depth.

Gains in feasibility appear to offset losses in reliability and validity when switching to a shorter measure like the ORS. Higher compliance rates were observed for the ORS in comparison with the OQ. Other anecdotal information collected by the first two authors support these results (i.e., everyday clinicians are far more inclined to use the ORS than the OQ).

One issue that appears to be strongly related to compliance is clinicians' perception of the usefulness of the measure to the therapeutic process. Many therapists see outcome measurement as an encumbrance to the process and an obstacle to forming alliances with clients. In consultation with a number of mental health agencies and third-party payers, the authors have repeatedly heard counselors report that outcome measurement is an "add-on" separate from actual clinical work and relevant only to management and other overseers. In such settings, measures that are easy to integrate into treatment and that have face validity 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. Obviously, no matter how reliable and valid the measure, if it is not used, the benefits of outcome management will not be realized, and the benefits are considerable as evidenced by recent studies demonstrating as much as a 150% improvement in clinical outcomes (Miller et al., 2003.).

All the problems typically associated with brief self-report tests (Boulet & Boss, 1991) apply to the 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 ORS does not measure nor is intended to identify clinical risk factors such as suicide or alcohol or drug use. Future research with more diverse clinical and nonclinical samples is underway and should further identify the strengths and weakness of the measure.

## NOTE

1. 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, and a school readiness program.

## REFERENCES

- Andrews, G. (1995). Best practices in implementing outcome management: More science, more art, worldwide. *Behavioral Healthcare Tomorrow*, 4(3), 19-24.
- Arneill, A. B., & Devlin, A. S. (2002). Perceived quality of care: The influence of the waiting room environment. *Journal of Environmental Psychology*, 22, 345-360.
- Boulet, J., & Boss, M. (1991). Reliability and validity of the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology*, 3(3), 433-437.
- 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 (Eds.), *The heart and soul of change* (pp. 389-406). Washington, DC: APA Press.
- Cohen, J., & Cohen, T. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.



- Dannon, P. N., Iancu, I., & Grunhaus, L. (2002). Psychoeducation in panic disorder patients: Effect of a self-information booklet in a randomized, masked-rater study. *Depression & Anxiety, 16*(2), 71–76.
- Duncan, B., & Miller, S. (2000). *The heroic client: Doing client directed, outcome informed therapy*. San Francisco: Jossey Bass.
- 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.
- 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.
- Grunhaus, L., Dolberg, O. T., Polak, D., & Dannon, P. N. (2002). Monitoring the response to rTMS in depression with visual analog scales. *Human Psychopharmacology Clinical & Experimental, 17*(7), 349–352.
- Humphreys, K. (1996). Clinical psychologists as psychotherapists: History, future, alternatives. *American Psychologist, 51*(3), 190–197.
- Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 19–71). New York: John Wiley.
- Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 143–189). New York: John Wiley.
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., et al. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 3*, 249–258.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, et al. (1996). *Administration and scoring manual for the OQ-45.2*. Stevenson, MD: American Professional Credentialing Services.
- Lambert, M. J., & Hills, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 72–113). New York: John Wiley.
- Lambert, M. J., Ogles, B., & Masters, K. (1992). Choosing outcome assessment devices: An organization and conceptual scheme. *Journal of Counseling and Development, 70*, 527–539.
- Lambert, M. J., Okiishi, J., Finch, A. E., & Johnson, L. D. (1998). Outcome assessment: From conceptualization to implementation. *Professional Psychology, 29*, 63–70.
- 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.
- 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.
- Miller, S. D., & Duncan, B. L. (2000). *The Outcome Rating Scale*. Chicago: Author.
- Miller, S., Duncan, B., Brown, J., Sorrell, R., & Chalk, M. (2003). *The Outcome Rating Scale and the improvement of effectiveness in a telephonic EAP setting*. Unpublished manuscript.
- Sanderson, W., Riley, W., & Eshun, S. (1997). Report of the working group on clinical services. *Journal of Clinical Psychology in Medical Settings, 4*(1), 3–12.
- Radbruch, L., Loick, G., Kiencke, P., Lindena, G., Sabatowski, R., Grond, S., et al. (1999). Validation of the German version of the Brief Pain Inventory. *Journal of Pain & Symptom Management, 18*(3), 180–187.
- 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.

**Acknowledgments.** The authors wish to thank Anne M. Allison, BA, who assisted with data collection and statistical analysis, and Karen Kinchin, MS, who helped with data management.

**Offprints.** Requests for offprints should be directed to Scott D. Miller, PhD, PO Box 578264, Chicago, IL 60657-8264. E-mail: scottdmiller@talkingcure.com

## APPENDIX

### Outcome Rating Scale (ORS)

Name _____	Age (Yrs): _____
ID# _____	Sex: M / F _____
Session # _____	Date: _____

---

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.

---

#### Overall:

(General sense of well-being)

|-----|

#### Individually:

(Personal well-being)

|-----|

#### Interpersonally:

(Family, close relationships)

|-----|

#### Socially:

(Work, School, Friendships)

|-----|

Institute for the Study of Therapeutic Change

www.talkingcure.com

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

Licensed for personal use only