

## Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures

ALISTAIR CAMPBELL<sup>1</sup> & SAMANTHA HEMSLEY<sup>2</sup>

<sup>1</sup>*School of Psychology, James Cook University, Townsville, Queensland and* <sup>2</sup>*Australian Institute of Health and Welfare, Canberra, Australian Capital Territory, Australia*

### Abstract

The validity and reliability of the Outcome Rating Scale (ORS) and the Session Rating Scale (SRS) were evaluated against existing longer measures, including the Outcome Questionnaire-45, Working Alliance Inventory, Depression Anxiety Stress Scale-21, Quality of Life Scale, Rosenberg Self-Esteem Scale and General Self-efficacy Scale. The measures were administered to patients referred for psychological services to a rural primary health-care service. Participants were recruited from both current and new patients of psychologists providing the service. Both the ORS and SRS demonstrated good reliability and concurrent validity with their longer alternatives. The ORS also evidenced significant correlations with measures of self-esteem, self-efficacy, and quality of life. The ORS and SRS offer benefits such as cost-effectiveness, brevity, simple administration, and easy interpretation of results in the measurement of clinical outcomes when compared to their longer counterparts. These results provide clear support for the adoption of brief outcome assessment measures in psychological practice.

**Keywords:** *Clinical/counselling psychology, discipline issues, Outcome Rating Scale, Session Rating Scale, theoretical and methodological issues*

Outcome assessment measures are ultimately intended to guide clinicians in tailoring treatment and to identify efficient treatment approaches (Smith, Fischer, Nordquist, Mosley, & Ledbetter, 1997). Many current outcome assessment measures lack essential components such as brevity, ease of administration, and content simplicity, all of which are vital if such measures are to be used on a session-by-session basis to enhance patient care. In recent years, two brief outcome assessment measures have been developed in an attempt to provide rapid but valid measurement. These are the Outcome Rating Scale (ORS) and the Session Rating Scale (SRS) (Miller & Duncan, 2000). Given the recency of their arrival in the psychotherapeutic arena, the reliability and validity of these measures has not adequately been tested. The present study was conducted in an attempt to evaluate the use of the ORS and SRS in psychological practice against other more established measures of outcome.

### *Practice-based evidence and outcome assessment measures*

Practice-based evidence requires that practitioners adopt a highly individualised service delivery plan for each patient, acknowledging the patient's goals for treatment, ideas about how change occurs, and view of an effective therapeutic relationship (Miller, Duncan, & Hubble, 2004). It supports the use of regular measures of patient progress through the gathering and dissemination of patient feedback and allows therapists to monitor their work in a systematic and ongoing fashion to ensure that it continually reflects the needs and treatment goals of the patient (Saggese, 2005).

Outcome assessment measures are recognised as an excellent method to operationalise practice-based evidence through the collection and communication of patient feedback. Utilisation of standardised measures that communicate ongoing outcomes allows practitioners to predict with a high degree of certainty the value of therapy and the continuity of

their services. The flexibility of such measures, however, also allows therapists to identify which patients are not responding to treatment and to adjust therapy accordingly (Saggese, 2005).

Outcomes measurement provides a number of benefits to a range of stakeholders at multiple levels. It provides practitioners with systematic needs assessment information and puts research into the hands of the clinician. This improves on the generalisability problems that characterise typical empirically validated treatments (Asay et al., 2002). Through the use of self-report outcome instruments, patients also benefit by actively monitoring their own progress both during and after treatment (Victoria's Mental Health Services, 2006). Finally, outcome measures simplify the interaction between external agencies and health-care professionals by providing hard data upon which decisions can be based regarding funding, insurance reimbursement, and the cost-effectiveness of psychotherapeutic services (Saggese, 2005).

There are problems, however, with the practicality of using such measures on a routine basis. Often, the methodological complexity, length of administration, and cost of outcome assessment measures render them infeasible because few practitioners have the time or resources to devote to the repeated administration, scoring and interpretation of their results (Miller, Duncan, Brown, Sparks, & Claud, 2003). Another major factor contributing towards the poor acceptance of outcome measures is the current focus on the use of outcome information. Much of the present literature discussing the advantages of outcome measurement focuses on economic decision-making rather than potential benefits to patients and clinicians (Stedman, Yellowlees, Mellisop, Clarke, & Crake, 1997). The Outcome Questionnaire-45 (OQ-45) and Working Alliance Inventory (WAI) are two examples of current outcome assessment measures that suffer many of the feasibility issues mentioned.

The OQ-45 is a 45-item self-report scale designed to assess three domains of functioning on a sessional basis. These include symptoms of psychological disturbance, interpersonal problems, and social role functioning (Lambert, Harmon, Slade, Whipple, & Hawkins, 2005; Wampold, 2001). This feedback allows clinicians to predict functioning at treatment termination and to adjust therapy accordingly. The OQ-45 is reported to have good reliability and validity across a number of settings and patient populations, and demonstrates adequate sensitivity to change (Lambert et al., 2005). But despite its excellent psychometric properties and widespread use, the OQ-45 is often rendered impractical due to the length of time needed to complete the

questionnaire, size of print, and content complexity (Miller et al., 2003).

The WAI is another self-report instrument designed to measure the quality of alliance between patient and therapist. Measures of process in psychotherapy are a necessary response to evidence that the quality of the patient-therapist relationship is a primary ingredient in psychotherapeutic effectiveness and is a reliable predictor of treatment outcome (Wampold, 2001). Despite the connection, however, between alliance and outcome, no measures have specifically been developed for day-to-day use. Like the OQ-45, the WAI demonstrates strong reliability and validity, although it lacks the essential component of utility. The length and complexity of the questionnaire makes it impracticable for everyday use due to the fact that it was largely developed for research purposes (Duncan et al., 2003).

Patients, clinicians, third-party payers, government agencies and researchers all benefit from the assessment of treatment outcome within the mental health field, but the utility of this information depends on the needs of the individual or group in question. For example, clinicians desire an instrument that does not significantly add to their own paperwork, is clinically relevant for a diverse patient population, and is easy to use, score, and interpret. Managed care administrators and sponsors share many of the same concerns but add factors such as cost, patient friendliness, and the existence of relevant benchmarks for evaluating patient change (Dunn, Burlingame, Walbridge, Smith, & Crum, 2005).

The use of brief outcome measures in clinical practice involves a simple methodology that can be easily implemented by most services. The simplicity of scoring procedures also leads to measurements that are easily interpreted. The cost of materials is low, which ensures that service providers are able to repeatedly administer the measures without significant cost. The implementation procedures of the scales are also usually simple, meaning that there is minimal training required in the administration, scoring, and interpretation of results. In addition, brief outcome assessment measures have obvious face validity and meet the minimal criteria of psychometric adequacy (Stedman et al., 1997). The ORS and SRS are two ultra-brief outcome assessment measures that have been developed in an attempt to meet these requirements.

The ORS was developed by Miller and Duncan (2000) as a brief alternative to the OQ-45 to assess change in patients following psychological intervention. Thus, it assesses the same three areas of functioning including individual, relational, and social, but does so in visual analogue format through a set of only four questions that take approximately 1 min to complete.

In order to determine whether or not the ORS was psychometrically able to act as an alternative measure of outcome, the reliability and validity of both the OQ-45 and ORS were assessed. Miller et al. (2003) assessed the validity and reliability of the ORS using both a non-clinical ( $n=86$ ) and clinical sample ( $n=435$ ). Participants in the non-clinical sample received four concurrent administrations of the ORS and OQ-45 over a period ranging from 10 days to 5 weeks. Participants in the clinical sample received only the ORS as part of standard treatment.

Results indicated that the ORS possessed a high degree of internal consistency for the non-clinical sample ( $\alpha=.93$ ) and compared favourably with that reported for the OQ-45. As would be expected from an ultra-brief measure, the test-retest reliability of the ORS (.66 at second administration) was lower than the OQ-45 (.83 at second administration) (Miller et al., 2003). Concurrent validity was also assessed on the non-clinical sample and results indicated moderately strong correlations between ORS items and OQ-45 subscales and total scores. In addition, an overall correlation of .59 was found between the ORS and OQ-45 total scores, demonstrating that the ORS was moderately related to the gold standard of self-report scales that the OQ-45 reflects.

One of the main reasons for the development of the ORS was to provide clinicians and patients with an outcome measurement tool that could be easily implemented on a routine basis within everyday clinical practice. Thus, to demonstrate the feasibility of the ORS, Miller et al. (2003) examined and compared compliance rates with the ORS and OQ-45 over a period of 12 months in a group of therapists ( $n=86$ ) practising at a community family service agency. A compliance rate of 89% was achieved for the ORS, whereas only a 25% compliance rate was achieved for the OQ-45. It was evident that the brevity, simple content structure, and scoring procedure of the ORS appealed to clinicians, who struggled to adopt the position of scientist-practitioner using more complex measures such as the OQ-45.

The SRS is another four-item visual analogue instrument that is based on encouraging patients to identify any alliance problems with their therapist so that the clinician may change to better fit patient expectations. This measure was originally developed by Johnson (1995) as a 10-item Likert-scaled instrument, but concerns regarding the time needed to complete the questionnaire were quick to surface among clinicians and patients. The SRS was developed as a brief alternative to the longer measure (Miller & Duncan, 2000).

Like the ORS, it was necessary to examine the reliability and validity of the SRS in order to

determine whether or not it could act as an alternative to instruments such as the WAI and Revised Helping Alliance Questionnaire (HAQ-II). Duncan et al. (2003) recruited participants from three clinical sites, including patients from an outpatient mental health counselling agency ( $n=81$ ), and therapists from the Family Therapy Associates ( $n=106$ ) and a community family service agency ( $n=50$ ). Participants from the outpatient mental health counselling agency completed six concurrent administrations of the SRS and HAQ-II to determine the reliability and validity of the measure.

Results indicated that Cronbach's alpha for the SRS (.88) compared favourably with that reported for the HAQ-II (.90). Test-retest reliability also demonstrated promising results (SRS  $r=.64$ , HAQ-II  $r=.63$ ). Concurrent validity calculations provided a correlation of .48 between the SRS and HAQ-II. All correlations between SRS items and total HAQ-II scores were within the range of .39 to .44 (Duncan et al., 2003). The feasibility of the SRS was also assessed by comparing the utilisation rate of the ultra-brief measure to that of the WAI among therapists ( $n=156$ ) from two clinical sites. The SRS had a 96% compliance rate while the WAI was used only 29% of the time. Based on this evidence the researchers described the SRS as an ultra-brief alternative for assessing global strength of alliance similar to that measured by longer, research-oriented alliance measures (Duncan et al., 2003).

#### Present Study

Despite early findings that both the ORS and SRS demonstrate solid reliability, validity, and feasibility by the researchers who developed these brief assessment tools, it remains to be seen whether these findings can be replicated. In addition, it is also an open question whether or not these brief outcome measures have the same psychometric properties when compared to other elaborate assessments of treatment outcome. The aim of the present study was to evaluate the validity of the ORS and SRS in psychological practice by comparing the outcome assessment data obtained from these measures with those from longer, more established measures. These included the OQ-45, the Quality of Life Scale (QOLS), the Depression Anxiety Stress Scale-21 (DASS-21), the General Self-Evaluation Scale (GSE), and the WAI.

#### Method

##### Participants

Data were collected from patients referred for psychological services to the North and West Queensland Primary Health Care (NWQPHC),

Division of General Practice. All current and any new patients thereafter for the period of the study were asked to take part in the research. A total of 65 participants were recruited for the study over a period of 7 months (54 female, 10 male, one unspecified). All participants were required to be aged  $\geq 18$  years. The mean age of the sample was 43 years ( $SD = 12.59$ ) with an age range of 18–62 years. The main presenting problem of each patient was identified from a list of 25 commonly occurring problems. Main presenting problem was identified based on clinical judgement rather than on formal diagnosis. The most common presenting problem was depression (30.8%), followed by anxiety (15.4%) and family and relationship issues (15.4%).

#### Procedure

A number of measures were used to evaluate the validity of the ORS and SRS in psychological practice. These include the OQ-45, QOLS, DASS-21, GSE, and WAI. These measures assess a number of domains including outcome, clinical presentation, symptom severity, therapeutic relationship, and quality of life.

**Outcome Questionnaire-45.** The OQ-45 consists of 45 self-report items covering a broad range of symptom and functioning domains, including subjective discomfort, interpersonal relationships, and social role performance. The underlying factor structure of the OQ-45 has been confirmed (Meuller, Lambert, & Burlingame, 1998) although the subscales are acknowledged to be highly inter-correlated (Umphress, 1997). Scores are generated for each subscale and a total score is obtained by summing the three subscale scores. Each item response is measured on a 5-point scale yielding a range of possible scores from 0 to 180 (Lambert et al., 2002). The OQ-45 provides criterion measures for the classification of patients into outcome groups including recovered, improved, deteriorated, and no change (Lambert, Okiishi, Finch, & Johnson, 1998). The OQ-45 has high reliability (.93) and evidence to suggest good concurrent and construct validity across a wide range of patient populations (Lambert & Hawkins, 2004). It is also sensitive to change over short periods in clinical samples while remaining stable in untreated individuals (Lambert et al., 2005).

**Quality of Life Scale.** The QOLS contains 16 items and is measured on a 7-point scale (7 = *delighted*, 1 = *terrible*) (Burckhardt & Anderson, 2003). The 16-item instrument is a valid measure of quality of life, and measures constructs distinct from health status and disease activity. The QOLS demonstrates strong

reliability estimates ( $\alpha = .82-.88$ ) and yields a score range from 16 to 112. The average score for healthy populations is 90 (Burckhardt, Archenholtz, & Bjelle, 1992).

**Depression Anxiety Stress Scale-21.** Like the original DASS created by Lovibond and Lovibond (1995), the DASS-21 consists of three subscales: depression, anxiety, and stress. Completion time, however, is significantly decreased by halving the number of questions in each subscale from 14 to 7. Previous research assessing the reliability of the DASS-21 found Cronbach's alpha coefficients of .94 for depression, .87 for anxiety, and .91 for stress. Concurrent validity was also assessed and results indicated moderate to strong correlations (Antony, Bieling, Cox, Enns, & Swinson, 1998). Scores are generated for each subscale and then doubled so that interpretations can be made based on the 42-item DASS scoring method (Lovibond & Lovibond, 1995). The factor structure of the DASS is well established (Antony et al., 1998; Norton, 2007).

**Rosenberg Self-Esteem Scale and General Perceived Self-Efficacy.** For ease of test administration both the Rosenberg Self-Esteem Scale (RSES) and General Perceived Self-Efficacy (GPSE) scale were combined to form the GSE, but data analysis was performed individually. Both the GPSE and RSES are scored on a 4-point Likert scale yielding total scores for each ranging from 10 to 40. The RSES is the most widely used self-esteem measure and has good reliability ( $\alpha = .88$ ) and validity estimates. A high score on the RSES indicates a high sense of self-esteem (Torrey, Mueser, Meltugo, & Drake, 2000). As yet, no normative data exist for the GPSE English-language version, although high internal consistency ratings ( $\alpha = .92-.93$ ) and validity estimates have been reported for the German version (Schwarzer, 1994). A high score on the GPSE indicates a greater sense of self-efficacy.

**Working Alliance Inventory-12.** The 12-item WAI is a short form of the original 36-item patient version and is scored on a 7-point Likert scale (1 = *does not correspond*, 7 = *corresponds exactly*) (Horvath & Greenberg, 1989). The four highest-loading items on the Task, Bond, and Goal subscales were used to form the WAI-Short, and each subscale demonstrates strong internal consistency estimates ( $\alpha = .90, .92,$  and  $.90$ , respectively) (Tracey & Kokotovic, 1989). Concurrent validity estimates suggest a strong relationship between WAI and WAI-12 subscale and total scores (Busseri & Tyler, 2003).

**Outcome Rating Scale.** The ORS is a four-item (overall, individually, interpersonally, socially) visual

analogue scale developed as a brief alternative to the OQ-45. It demonstrates strong reliability estimates ( $\alpha = .87-.96$ ) and moderate correlations between the ORS items and OQ-45 subscale and total scores (ORS total-OQ-45 total:  $r = .59$ ). This correlation meets expectations given that 45 items were reduced to four (Miller et al., 2003). A copy of the ORS can be found in Appendix A. Respondents fill in the ORS by marking agreement with four statements on a visual analogue scale anchored at one end by the word *Low* and at the other end by the word *High*. The respondent is asked to rate how they have been feeling about their general wellbeing, personal wellbeing, family relationships, and social relationships since the last contact. This yields four separate scores between 0 and 100 using a millimetre for scale measurement.

**Session Rating Scale.** The SRS is a four-item (relationship, goals and topics, approach and method, overall) visual analogue scale designed specifically for everyday clinical use. It demonstrates good reliability estimates across a range of patient populations (.88) and is recognised as a valid measure of therapeutic alliance. The SRS is positively correlated with measures of outcome such as the ORS (.29,  $p < .01$ ), indicating that the SRS functions in much the same way as other alliance measures (Duncan et al., 2003). A copy of the SRS can be found in Appendix B. Respondents fill in the SRS by marking agreement with four statements on a visual analogue scale anchored at each end by statements about how they related to the therapist. The respondent is asked to rate the relationship, goals/method, approach/method and overall experience of the therapy session. This yields four separate scores between 0 and 100 using a millimetre for scale measurement.

The data collection phase of the present research was conducted over a period of 7 months and took only one consultation session to complete per participant. The participants involved in this research were recruited through NWQPHC psychologists providing services throughout North Queensland, who voluntarily agreed to assist in the test administration phase of this research. The participants were informed of the research and its purposes at the end of their first consultation session during the research period, and those who gave consent subsequently completed the required questionnaires. This research was fully supported by the NWQPHC Division of GPs.

The present data were obtained during the first administration of the ORS and SRS with the primary practice patients. This occurred at the end of the first or second session of therapy for new patients and at the closest session to the start of a study for existing patients. The psychologists involved were generally

eclectic in their approach and utilised a mix of cognitive behavioural therapy, psychoeducation, and interpersonal psychotherapy.

#### Statistical analysis

Basic statistical analyses were conducted using SPSS Version 12.0 (SPSS, Chicago, IL, USA). The internal consistency of the measures was determined using Cronbach's coefficient alpha, and a series of bivariate correlations was performed to assess the concurrent validity of the questionnaires.

## Results

#### Normative data

In line with previous research, OQ-45 scores were normally distributed (Shapiro-Wilks,  $p > .05$ ). Shapiro-Wilks tests, however, showed that the distribution of ORS scores deviated significantly from a normal distribution. In particular, the "Overall" and "Individually" items appeared to have bimodal distributions, peaking at both the low and high ends of the scale. The distributions of SRS and WAI scores were negatively skewed ( $M = 3.64$  and  $74.34$ , respectively); this was expected, however, given the nature of the scales. The parametric analyses of the non-normal data were compared to non-parametric analyses and the violations of normality did not have a significant effect. Consequently the parametric analyses were reported throughout.

#### Reliability

Internal consistency was evaluated for the subscale and total score of each measure. Table I displays Cronbach's coefficient alphas for each measure and shows that all measures were highly reliable ( $\alpha = .89-.95$ ). The reliability coefficients for the ORS and SRS were particularly impressive given the

Table I. Cronbach's alpha reliability coefficients

Instrument	$\alpha$
ORS	.90
SRS	.93
OQ-45	.95
WAI	.91
DASS-21	.95
QOLS	.90
RSES	.91
GPSE	.89

Note: DASS-21 = Depression Anxiety Stress Scale-21; GPSE = General Perceived Self-Efficacy Scale; OQ-45 = Outcome Questionnaire-45; ORS = Outcome Rating Scale; QOLS = Quality of Life Scale; RSES = Rosenberg Self-Esteem Scale; SRS = Session Rating Scale; WAI = Working Alliance Inventory.

Table II. Outcome rating scale and session rating scale intercorrelations

	1	2	3	4	5	6	7	8
1 ORS: Overall	1.00	.96	.61	.70	.07	.24	.16	.17
2 ORS: Individually		1.00	.63	.70	.09	.25	.18	.21
3 ORS: Interpersonally			1.00	.58	.13	.19	.22	.18
4 ORS: Socially				1.00	.03	.12	.12	.15
5 SRS: Relationship					1.00	.74	.79	.76
6 SRS: Goals and Topics						1.00	.80	.80
7 SRS: Approach or Method							1.00	.86
8 SRS: Overall								1.00

Notes: ORS = Outcome Rating Scale; SRS = Session Rating Scale.  
**Bold**,  $p < .05$ .

small number of items in each scale ( $\alpha = .90$  and  $.93$ , respectively).

Inter-item correlations were calculated to examine the extent to which each item represented the total scale. The inter-item correlations for the ORS and SRS are displayed in Table II. An inspection of the results indicated moderate to strong correlations between the four ORS items ( $r = .58-.97$ ). In particular, a strong correlation was found between "Overall" and "Individually" ( $r = .97, p < .01$ ). The SRS displayed a consistent pattern of strong inter-item correlations ( $r = .74-.86, p < .01$ ). The ORS and SRS items were not strongly correlated and, apart from a weak correlation between "Goals and Topics" on the SRS and individual wellbeing on the ORS, none of the correlations was significant.

Item-total correlations were also calculated for each of the other measures. Three items on the OQ-45 did not appear to accurately represent the corresponding subscale. Item 11 ("After heavy drinking, I need a drink the next morning to get going": Symptoms of Distress) had a corrected item-total correlation of  $-.02$ ; item 26 ("I feel annoyed by people who criticise my drinking/drug use": Interpersonal Relationships) had a corrected item-total correlation of  $-.08$ ; and item 32 ("I have trouble at work/school because of drinking/drug use": Social Role) had a corrected item-total correlation of  $-.001$ . These items suggested that the drinking/drug use items were not stable in this sample.

### Validity

Concurrent validity was computed using Pearson product-moment correlations (Table III). In support of previous research, moderate to strong correlations were found between the ORS items and OQ-45 subscale and total scores. In particular, strong correlations were found between OQ-45 Symptoms of Distress and ORS Overall ( $r = -.75, p < .01$ ) and Individually ( $r = -.74, p < .01$ ), but only moderate correlations were found between all other subscales. The ORS items were significantly correlated with the DASS scales. The strongest relationships were

between the DASS Depression and Stress scales and the Overall and Individually wellbeing items on the ORS. The DASS Anxiety scores were moderately correlated with all but the Interpersonal wellbeing item on the ORS, which was not significant. The ORS items also had moderate to strong correlations with self-esteem (RSES) (range:  $.46-.67$ ) and moderate correlations with self-efficacy (GPSE) (range:  $.36-.53$ ). Strong correlations were found between the ORS items and quality of life (QOLS) (range:  $.49-.74$ ).

The SRS had significant and consistently moderate correlations with the WAI. All correlations between SRS items and WAI subscales were within the range of  $.37-.63$ . In addition, the correlations did not identify any specific relationships between the SRS items and WAI subscales, such as SRS Relationship and WAI Bond. Instead, relatively equal correlations were found between all subscales. There were some weak but significant correlations between the SRS and the OQ-45 Interpersonal Relationships subscale (range:  $-.28$  to  $-.36$ ) and the Total score (range:  $-.27$  to  $-.30$ ) but there were no significant correlations with the DASS scales or with quality of life, and self-efficacy. One weak but significant correlation was found between the SRS Overall item and self-esteem ( $r = .24, p < .05$ ).

### Discussion

The strong inter-item correlations for both the ORS and SRS suggested that a single dimension underlying each scale could be an interpretable summary. Although it is not common practice, it would be reasonable to summarise the item scores into a single score to globally represent outcome and session relationship, respectively. These findings support previous research, suggesting that the ORS is best thought of as a global measure of distress (Miller et al., 2003), and the SRS a global measure of alliance (Miller et al., 2003). The very strong correlations between the ORS items of Overall wellbeing and Individually wellbeing raises the possibility that these two items are measuring the

Table III. ORS-SRS subscale and total correlations with validating measures

	OQ-45		WAI			DASS-21			QOLS		RSES		GPSE	
	OQ-45 Symptoms of Distress	OQ-45 Interpersonal Relationships	OQ-45 Social Role	OQ-45 Total	WAI Goal	WAI Task	WAI Bond	WAI Total	DASS Dep.	DASS Anxiety	DASS Stress	RSES Total	RSES Total	GPSE Total
ORS Overall	-.75*	-.56*	-.57*	-.76*	.14	.19	.80	.19	-.76*	-.54*	-.67*	.67*	.67*	.51*
Individually	-.74*	-.55*	-.55*	-.74*	.17	.17	.16	.18	-.74*	-.52*	-.66*	.67*	.67*	.53*
Interpersonal	-.36*	-.54*	-.24	-.42*	.12	.11	.29*	.19	-.40*	-.22	-.34*	.46*	.46*	.36*
Socially	-.63*	-.56*	-.55*	-.67*	.08	.13	.16	.14	-.59*	-.34*	-.43*	.53*	.53*	.45*
Total	-.71*	-.63*	-.55*	-.74*	.15	.17	.23	.20	-.71*	-.46*	-.60*	.66*	.66*	.53*
SRS														
Relationship	-.16	-.28*	-.14	-.20	.46*	.42*	.37*	.47*	-.07	-.06	0.09	0.17	0.17	0.21
Goals and Topics	-.22	-.33*	-.19	-.27*	.60*	.60*	.52*	.65*	-.17	-.16	-.04	.22	.22	.17
Approach or Method	-.23	-.36*	-.14	-.27*	.63*	.62*	.56*	.68*	-.13	-.07	.06	.19	.19	.19
Overall	-.28*	-.32*	-.14	-.30*	.62*	.62*	.59*	.58*	-.19	-.14	.02	.24*	.24*	.29
Total	-.24	-.34*	-.17	-.28*	.62*	.62*	.49*	.63*	-.15	-.12	.04	.22	.22	.22

Note: DASS-21 = Depression Anxiety Stress Scale-21; GPSE = General Perceived Self-Efficacy Scale; OQ-45 = Outcome Questionnaire-45; ORS = Outcome Rating Scale; QOLS = Quality of Life Scale; RSES = Rosenberg Self-Esteem Scale; SRS = Session Rating Scale; WAI = Working Alliance Inventory.

same underlying construct. It is possible that the four-item ORS could be reduced to three items with no real loss of information. This may further enhance the clinical utility of the measure because it would be a very brief scale.

The clinical utility of the ORS is further supported by the strong positive correlations between it and the OQ45 and the DASS. This indicates that the ORS is able to provide rapid and valid information about patient functioning and wellbeing compared to these longer alternatives. The relatively stronger correlations with the DASS depression and DASS stress scales suggests that it may be more sensitive to these than to anxiety but this would need further clinical elaboration. The correlations between the ORS and scales of quality of life, self-esteem, and self-efficacy indicate that it is evaluating more than just symptom distress. This also supports the clinical utility of the ORS because the scores on the ORS appear to be tapping similar dimensions to the OQ-45, which is also generally well correlated with measures such as these. It should be noted, however, that there is a necessary cost to utilising ultra-brief outcome measures in terms of the loss of clinical information, which is traded for utility and brevity.

The relationship between the SRS and the WAI was generally encouraging but nowhere near as clear as it was with the ORS. This could be due to the nature of the scale but is also likely to do with difficulties with conceptualising the concept of therapeutic alliance and relationship. The correlations between the SRS and the WAI support the idea that the SRS is measuring a construct of therapeutic alliance, but the lack of specificity between the subscales suggests that the SRS may be simply tapping a general alliance construct. Nonetheless, the SRS must still be considered a good candidate for the pragmatic measurement of alliance because it is so much easier and quicker to use than the WAI and other existing measures of the therapeutic relationship.

The present small sample size limits the conclusions that we can make in relation to the ORS and the SRS. Nonetheless, our results are similar to those that have already been published and allow a conclusion that both scales have a high degree of clinical utility for evaluating outcome on the basis of more than only symptom presentation. Another limitation to the study was that the subjects were drawn from a primary care population, which is arguably different from a psychiatric clinical population in terms of both severity and the likelihood of formal DSM diagnoses. It would be useful to see an evaluation of ultra-brief outcome measures in other clinical populations covering the range of diagnostic conditions and severity. An additional direction for further research should also include confirmatory factor analyses on the various measures to further

explore the psychometric properties of existing measures as well as the ultra-brief measures.

Brief-outcome assessment measures, such as the ORS and SRS, should be carefully considered when developing measures of outcome in real-world settings (Campbell, 2002). One of the major blocks to good research in real-world settings is the compliance of clinicians with long and detailed measurement protocols. The rate of compliance with the ultra-brief measures ensures that much more information about the outcome of therapy and the therapeutic relationship can be collected than when more research-oriented measures are utilised. Compliance is a significant consideration in real-world research, which is required to establish an evidence base for the effectiveness, as opposed to simple efficacy (Campbell, 2005), of therapeutic activity.

## References

- Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress scales in clinical groups and a community sample. *Psychological Assessment, 10*, 176–181.
- Asay, T. P., Lambert, M. J., Gregersen, A. T., & Goates, M. K. (2002). Using patient-focused research in evaluating treatment outcome in private practice [electronic version]. *Journal of Clinical Psychology, 58*, 1213–1225.
- Burckhardt, C. S., & Anderson, K. L. (2003). The quality of life scale (QOLS): Reliability, validity and utilisation [electronic version]. *Health and Quality of Life Outcomes, 1*(60).
- Burckhardt, C. S., Archenholtz, B., & Bjelle, A. (1992). Measuring the quality of life of women with rheumatoid arthritis or systemic lupus erythematosus: A Swedish version of the quality of life scale (QOLS). *Scandinavian Journal of Rheumatology, 21*, 190–195.
- Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory: Short form. *Psychological Assessment, 15*, 193–197.
- Campbell, A. (2002). Evidence based practice: Is it good for you? *Australian and New Zealand Journal of Family Therapy, 23*, 215–217.
- Campbell, A. (2005, February). *Evidence based intervention in Australia: Where is the debate?* Paper presented at the Australian Counselling and Supervision Conference, Brisbane.
- Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., & Brown, J. (2003). The Session Rating Scale: Preliminary psychometric properties of a working alliance measure [electronic version]. *Journal of Brief Therapy, 3*, 3.
- Dunn, T. W., Burlingame, G. M., Walbridge, M., Smith, J., & Crum, M. J. (2005). Outcome assessment for children and adolescents: Psychometric validation of the Youth Outcome Questionnaire 30.1. *Clinical Psychology and Psychotherapy, 12*, 388–401.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counselling Psychology, 36*, 223–233.
- Johnson, L. D. (1995). *Psychotherapy in the age of accountability*. New York: Wiley Norton & Company.
- Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61*, 165–174.
- Lambert, M. J., & Hawkins, E. J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*, 492–499.
- Lambert, M. J., Okiishi, J. C., Finch, A. E., & Johnson, L. D. (1998). Outcome assessment: From conceptualisation to implementation. *Professional Psychology: Research and Practice, 29*, 63–70.
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., & Nielsen, S. L. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy, 9*, 91–103.
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck depression and anxiety scales. *Behaviour Research and Therapy, 33*, 335–343.
- Meuller, R., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70*, 248–262.
- Miller, S. D., & Duncan, B. L., & (2000). *The Outcome and Session Rating Scales: Administration and scoring manuals*. Retrieved 10 April 2006, from <http://www.talkingcure.com>
- 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 brief visual analog measure. *Journal of Brief Therapy, 2*, 91–100.
- Miller, S. D., Duncan, B. L., & Hubble, M. A. (2004). Beyond integration: The triumph of outcome over process in clinical practice [electronic version]. *Psychotherapy in Australia, 10*(2).
- Norton, P. J. (2007). Depression Anxiety and Stress scales (DASS): Psychometric analysis across four racial groups. *Anxiety, Stress, and Coping: An International Journal, 20*, 253–265.
- Saggese, M. L. (2005). Maximising treatment effectiveness in clinical practice: An outcome-informed, collaborative approach. *Families in Society: The Journal of Contemporary Social Services, 86*, 558–564.
- Schwarzer, R. (1994). Optimism, vulnerability, and self-beliefs as health-related cognitions: A systematic overview. *Psychology and Health, 9*, 161–180.
- Smith, G. R., Fischer, E. P., Nordquist, C. R., Mosley, C. L., & Ledbetter, N. S. (1997). Implementing outcomes management systems in mental health settings. *Psychiatric Services, 48*, 364–368.
- SPSS for Windows, Rel. 12.00. (2001). Chicago: SPSS Inc.
- Stedman, T., Yellowlees, P., Mellsop, G., Clarke, R., & Crake, S. (1997). *Measuring consumer outcomes in mental health*. Retrieved 28 January 2006, from <http://www.mhnocc.org>
- Torrey, W. C., Mueser, K. T., Meltugo, G. H., & Drake, R. E. (2000). Self-esteem as an outcome measure in studies of vocational rehabilitation for adults with severe mental illness. *Psychiatric Services, 51*, 229–233.
- Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 207–210.
- Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the Outcome Questionnaire. *Journal of Psychoeducational Assessment, 15*, 40–55.
- Victoria's Mental Health Services. (2006). *About outcome measurement, 2006*. Retrieved from <http://www.health.vic.gov.au/mentalhealth/outcomes/about.htm>
- Wampold, B. E. (2001). *The great psychotherapy debate*. New Jersey: Lawrence Erlbaum Associates.

## Appendix A

### Outcome Rating Scale

Looking back over the last week (or since your last visit), 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 represent high levels.

	<b>Overall:</b> (General sense of well-being)	
Low	-----	High
	<b>Individually:</b> (Personal well-being)	
Low	-----	High
	<b>Interpersonally:</b> Family, close relationships)	
Low	-----	High
	<b>Socially:</b> (Work, school, friendships)	
Low	-----	High

## Appendix B

### Session Rating Scale

Please rate today's session by placing a hash mark on the line nearest to the description that best fits your experience.

	<b>Relationship:</b>	
I did not feel heard, understood and respected	-----	I felt heard, understood and respected
	<b>Goals and Topics:</b>	
We did not work on or talk about what I wanted to work on or talk about	-----	We worked on and talked about what I wanted to work on or talk about.
	<b>Approach or Method:</b>	
The therapist's approach is not a good fit for me	-----	The therapist's approach is a good fit for me
	<b>Overall:</b>	
There was something missing in the session today	-----	Overall, today's session was right for me