

The Outcome and Session Rating Scales: A Cross-Cultural Examination of the Psychometric Properties of the Dutch Translation

Anton Hafkenscheid

Sinai Centre, Jewish Mental Health Services
Amersfoort, The Netherlands

Barry L. Duncan

Heart and Soul of Change Project
Tamarac, FL

Scott D. Miller

International Center for Clinical Excellence
Chicago, IL

Abstract

Continuous client feedback offers great promise for improving treatment outcomes. Because of their feasibility for everyday clinical use, the Outcome Rating Scale (ORS) and the Session Rating Scale (SRS) enjoy popularity in the US and have been translated in various languages, including Dutch. The present study investigated the psychometric properties of the Dutch translations of ORS and SRS and examined the cross-cultural differences between Dutch and US data. Results were largely consistent with the American studies, although some differences were found. This preliminary study of the Dutch translations supports the use of the ORS and SRS cross-culturally but additional research of other translations is needed.

Keywords: Outcome Rating Scale, Session Rating Scale, reliability, validity, cross-cultural, replication

This study was made possible by the Dutch Jewish Mental Health Services, Amersfoort, Amstelveen, the Netherlands. Special thanks are due to Mrs. Addie Cijssouw for secretarial assistance and to the clients and therapists who participated in this study. The consensus translations of the Dutch ORS and SRS were produced by Dotti Been, Sjouk de Boer, Albert Boon, Peter Breukers, Marc Crouzen, Pico Teune, and Joop Wolff, in collaboration with the first author.

Correspondence concerning this article should be addressed to Anton Hafkenscheid Sinai Centre, Jewish Mental Health Services, Outpatient Clinic, Berkenweg 7, 3818 LA Amersfoort, The Netherlands. Email: a.hafkenscheid@sinaicentrum.nl

It is often reported that the average treated person is better off than approximately 80% of the untreated sample ((Duncan, Miller, Wampold, & Hubble, 2010; Lambert & Ogles, 2004; translating to an effect size (ES) of about 0.8. Improvements, in other word, achieved by clients receiving “active” treatment typically surpass gains made in those who get a placebo (“inert” treatment) or go untreated. Generally, the effects not only reach statistical significance, but are also clinically meaningful (Lambert & Ogles, 2004). In spite of this encouraging conclusion, a substantial minority of clients still remains at risk of treatment failure or drop out. These clients do not benefit from psychotherapy or may even be harmed because of, for instance, inept application of treatment, negative attitudes of psychotherapists or poor combinations of technique and client problem (Lambert & Ogles, 2004).

Ideally, therapists should become aware as early as possible of problems in the therapeutic alliance or the absence of substantial progress, to prevent ultimate treatment failure. For different reasons, this ideal situation is far from commonplace. Unfortunately, therapists are poor at predicting client deterioration or drop out (Hannan et al., 2005). Thus, the development and implementation of reliable and clinically feasible instruments that assist therapists and clients in systematically monitoring the therapeutic alliance and treatment change is pertinent. Such measures aim to detect subtle drifts in client engagement in treatment or negative shifts in the client’s daily functioning.

Several measuring devices for the session-by session monitoring of treatment process and progress have been developed by researchers and clinicians from different therapeutic orientations (see Lambert, 2010 for a full discussion): the Outcome Questionnaire (OQ-45; de Jong et al., 2007; Lambert et al, 1996), the Clinical Outcomes and Routine Evaluation Outcome Measure (CORE-OM; Evans et al., 2002), the Session Evaluation Questionnaire (SEQ; Hafkenscheid, 2009a; Stiles et al., 1994), the Post Session Questionnaire (PSQ; Samstag, Batchelder, Muran, Safran, & Winston, 1998) and the combination of the Outcome Rating Scale (ORS) and Session Rating Scale (SRS; Duncan, 2010; Duncan & Sparks, 2010; Duncan et al., 2003; Miller & Duncan, 2004; Miller, Duncan, Brown, Sparks & Claud, 2003; Bringhurst, Watson, Miller, & Duncan, 2006). The ORS and SRS are the briefest, most generic, and most user-friendly monitoring instruments. Both measures are the core of the *Partners for Change Outcome Management System* (PCOMS), a client-based outcome feedback approach (Anker, Duncan, & Sparks, 2009; Duncan, Miller & Sparks, 2004; Miller, Duncan, Sorrell & Brown, 2005; Reese, Norsworthy, & Rowlands, in 2009).

Encouraging preliminary psychometric results were obtained with the American ORS (Bringhurst et al., 2006; Miller et al., 2003) and SRS (Duncan et al., 2003). Since the instruments were introduced in the US, the ORS and SRS have been translated into different languages: Finnish, French, German, Greek, Hebrew, Norwegian, Slovak, Spanish, Swedish and Dutch. For cross-cultural comparisons, translations of existing measures are preferable to the development of idiosyncratic, language-specific instruments. Non-English ORS and SRS versions should not only have the same format and item content as the original American version, but should also have similar psychometric qualities. No psychometric data regarding non-English ORS and SRS versions have been published so far, with the exception of the Norway Couple Feedback Project (Anker et al., 2009). The aim of the present study is to report psychometric findings with the Dutch ORS and SRS and compare the results with the ORS and SRS American studies.

Methods

Instruments

Outcome Rating Scale (ORS) and Session Rating Scale (SRS). The ORS and the SRS each consist of four items in visual analogue format. For each item, the respondent is instructed to place a

mark on a 10-cm line. The more the mark is placed to the left, the more negative the client's feels about her/his functioning in everyday life over the last week (ORS) or about that particular treatment session (SRS); the more a mark is placed to the right, the more positive the client's feeling. The ORS items cover three areas of client functioning: individually (personal well-being), interpersonally (family, close relationships) and socially (work, school, friendships). Three of the four SRS items cover the main elements of the therapeutic alliance: the relationship (on a continuum from "I did not feel heard, understood, and respected" to "I did feel heard, understood, and respected"), goals and topics (on a continuum from "We did not work on or talk about what I wanted to work on and talk about" to "We did work on or talk about what I wanted to work on and talk about") and approach or method (on a continuum from "The therapist's approach is not a good fit for me" to "The therapist's approach is a good fit for me"). The fourth item of both the ORS and the SRS requires the client to globally evaluate her/his daily functioning (ORS) and the treatment session (SRS). The ORS and SRS are scored by simply summing the marks made by the client measured to the nearest millimeter on each of the four lines¹.

A group of eight psychologists and psychotherapists from different parts of the Netherlands produced careful consensus translations of the English ORS and SRS. The only substantial difference between the American and Dutch versions of the instruments is the use of a question format (e.g., "How was your personal well-being last week?" instead of "Personal well-being") for the Dutch ORS. This question format was adopted from the child version of the American ORS (Duncan et al., 2006). The ultimate ORS and SRS translations were approved by the American owners of the instruments (Miller, personal communication, November 13, 2006).

Therapist Satisfaction Scale (TSS). The TSS (Tracey, 1992) is a therapist-based rating scale, containing seven items. Sample items are "I do not expect to be of any help to this client" or "I enjoy working with this client." Items are responded to by the therapist using a 7-point Likert-type scale, ranging from 1 (very strongly disagree) to 7 (very strongly agree). Scores on negative statements are reversed, yielding a one-dimensional total score, with higher scores denoting more therapist satisfaction. The TSS was translated in Dutch with permission of the author (Tracey, personal communication, June 21, 2005).

Participants

A heterogeneous convenience sample of 126 Dutch speaking clients referred to the outpatient clinic of the Sinai Centre (Jewish Mental Health Services, Amersfoort, the Netherlands) for treatment between 2006 and 2009 participated in the study. Age: $m = 51.9$ years ($sd = 14.1$; range: 19-81). Gender: 62 (49.2 %) female clients, 64 (50.8 %) male clients. Most of the participating patients were descendants of World War II (holocaust victims or survivors of the Japanese concentration camps, spouses and children of survivors), (Dutch speaking) refugees or young Dutch veterans of (United Nations or NATO) Balkan, Middle East or Central Asia peace missions.

Procedure

A total of 1005 ORS and SRS forms were completed by a cohort of 126 clients, each treated by one of 18 therapists. The number of forms completed by clients ranged from 3 to 10, with an average of 8 administrations per client. One therapist (the first author; a clinical psychologist and senior psychotherapist), treated 75 clients. The other 51 clients were treated by one of the other 17

¹ The ORS and SRS are free for individual use and available at www.heartandsoulofchange.com and www.scottdmiller.com.

therapists. Therapists of the 32 clients were 6 junior psychologists; therapists of the remaining 19 clients were senior therapists of various disciplines (social work, psychiatric nurses, drama therapists, and senior psychotherapists).

All therapists completed the 7-item Dutch version of the TSS after each session. Clients and therapists were blind to each other's ratings. Therapists' TSS ratings were used to estimate convergent validity, by correlating these ratings with ORS and SRS scores respectively. ORS and SRS item responses were measured by an independent secretary, who entered all ORS, SRS and TSS data into SPSS. No intermediate feedback was provided to clients or therapists.

Unselected client group. The 75 clients treated by the first therapist were unselected in the sense that all clients referred to him were invited and were willing to participate.

Selected client group. The 51 clients treated by the other 17 therapists were invited by their therapist to participate in a pilot study about the Dutch ORS and SRS. No inclusion or exclusion criteria were formulated. Therapists simply decided which clients would be invited to participate.

The following psychometric aspects of the ORS and SRS were examined: normative data, inter-correlations between both measures, homogeneity (internal consistencies) of the ORS and SRS, stability (test-retest reliability) and convergence of client ratings on the ORS and SRS with therapists' satisfaction (TSS global scores). Furthermore, predictive validity of the alliance measure (with the SRS as the predictor variable) for outcome status (with the ORS scores as the criterion variable) was determined in a subgroup of clients fulfilling two conditions: (a) completing enough ORS and SRS forms to allow robust estimates of predictive validity and (b) similar with respect to the number of ORS and SRS administrations completed, to control for differences in the amount of treatment received between the prediction and criterion measurement points. A subgroup of 68 clients completing 10 ORS and SRS forms fulfilled both conditions. In addition, cross cultural comparisons between the Dutch and American data normative were made regarding normative data, internal consistency and test-retest reliability.

Results

Preliminary analyses

To maximize sample size, the 126 clients would preferably be treated as a single sample. However, sampling differences between the unselected and selected client groups may be reflected in client's ORS and SRS ratings. Combining both groups is only justified if clients in the unselected group produce ORS and SRS scores comparable to those produced by clients in the selected group. Differences between global ORS and SRS scores of both client groups were tested at each of the ten administrations of the instruments, using *t*-tests for independent samples (two-tailed). For the ORS, none of the mean differences reached significance (first three administrations: $0.01 \leq t \leq 0.86$; $0.65 \leq p \leq 0.99$; $df = 124$; p -values for the other seven administrations: $0.31 \leq p \leq 0.94$).

For the SRS, a significant difference was found only at the ninth administration ($t = 1.96$; $p = 0.05$; $df = 75$), with higher SRS scores for the unselected client group. None of the other differences reached statistical significance (first three administrations: $0.01 \leq t \leq 0.86$; $0.39 \leq p \leq 0.99$; $df = 124$; p values for the remaining six administrations: $0.41 \leq p \leq 0.77$). These results justify the combination of ORS and SRS data from each of the client groups into one aggregated sample.

Normative data

Table 1 displays the means and standard deviations of the ORS and SRS for the first three administrations of the instruments, and also for all 1005 administrations of the scales.

Table 1: Summary statistics ORS and SRS

| | Administration | | | |
|----------------------|-----------------------|-----------------------|-----------------------|--|
| | <u>1st</u> | <u>2nd</u> | <u>3rd</u> | <u>1st to 10th</u> |
| m _{ORS} = | 19.3 | 20.7 | 20.5 | 21.2 |
| sd _{ORS} = | 8.2 | 8.9 | 9.0 | 9.3 |
| Mdn _{ORS} = | 20 | 21 | 21 | 21 |
| ≤ 25 | 75.4 % | 68.3 % | 69 % | 65.2 % |
| m _{SRS} = | 32.5 | 31.9 | 32.1 | 32.4 |
| sd _{SRS} = | 5.7 | 6.2 | 5.9 | 5.9 |
| Mdn _{SRS} = | 34 | 33 | 34 | 34 |
| ≤ 36 | 69 % | 69 % | 69.8 % | 67.8 % |
| n = | 126 | 126 | 126 | 1005 |

Across all 1005 ratings, the ORS mean global score is 21.2 (sd = 9.3; range: 0.5 to 39.3) and the SRS mean global score is 32.4 (sd= 5.9; range: 7.3 to 40). The distributions obtained for the global ORS and SRS scores across all 1005 ratings are presented in Figure 1 (ORS) and Figure 2 (SRS). Figure 1 shows that global ORS scores are fairly symmetrically distributed. In contrast, SRS scores appear to be highly skewed to the right (Figure 2), denoting that positive ratings of the therapeutic alliance are prevalent.

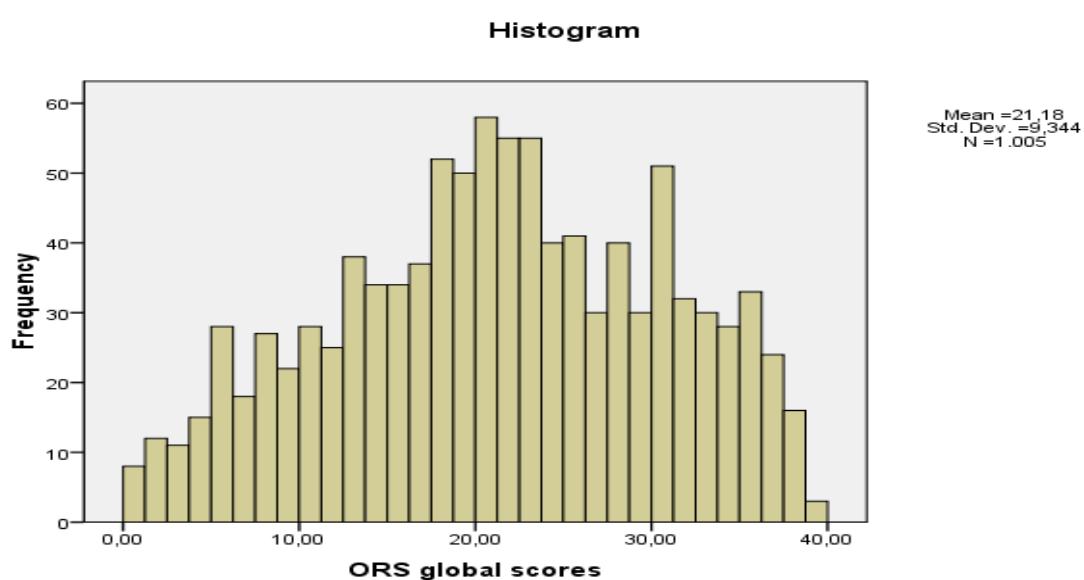
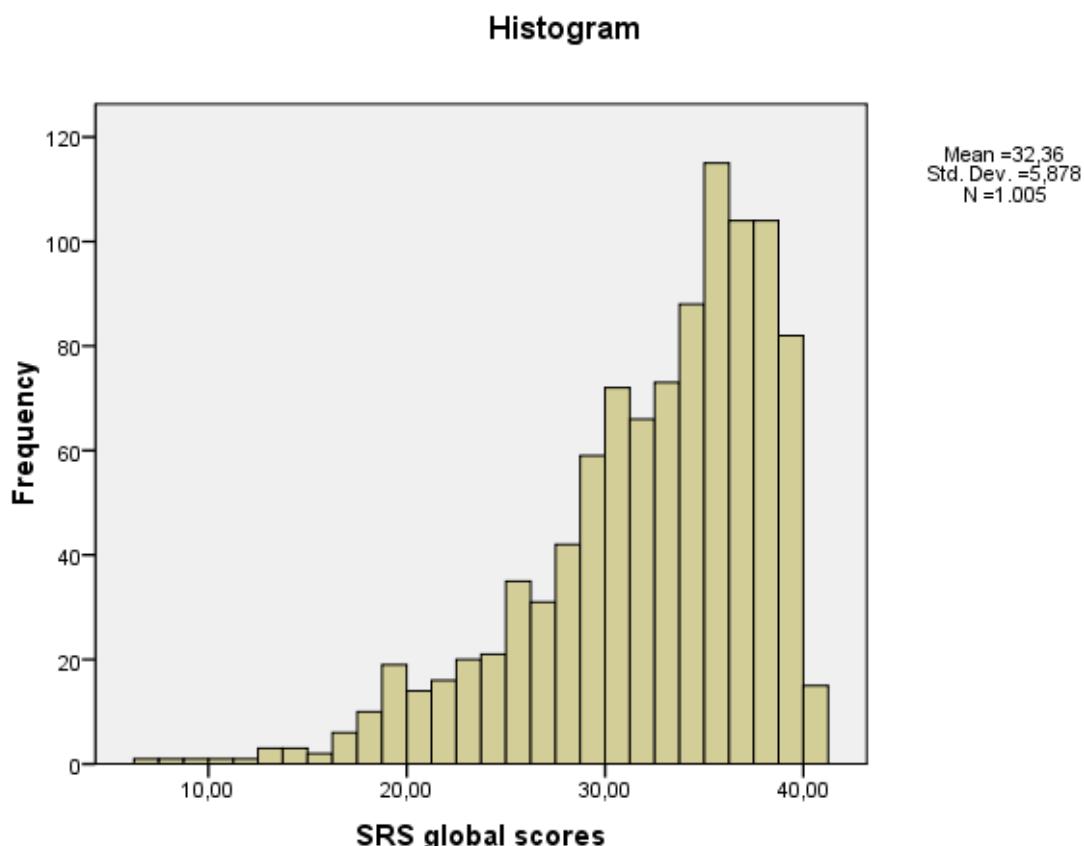
Figure 1: Distribution of global ORS scores (1005 administrations)

Figure 2: Distribution of global SRS scores (1005 administrations)



Comparison with American results. The mean initial ORS score in this Dutch sample (19.3) is similar to the mean ORS score (19.6) found in a clinical sample by Miller and Duncan (2004; Table 3, p. 9). Also, the standard deviations of the Dutch and American ($sd = 8.7$) samples are comparable. For the SRS, no descriptive sample statistics are reported in the American SRS publications, precluding comparisons with the American results.

In the American studies; the clinical cut-off for the ORS has been defined at a score of 25. No percentages below or above this cut-off score are reported. The clinical cut-off for the American SRS has been determined at a score of 36, with slightly fewer than 24% of cases scoring lower than this cut-off score (Miller & Duncan, 2004, p. 14). As Table 1 shows, about two third to three quarter of the Dutch ORS and SRS scores are below the American cut-offs.

Correlations between ORS and SRS

Table 2 : Inter-correlations ORS and SRS

| Administration | <u>1st</u> | <u>2nd</u> | <u>3rd</u> | <u>1st to 10th</u> |
|------------------|-----------------------|-----------------------|-----------------------|--|
| <i>r</i> ORS SRS | .11 | .33** | .25** | .28** |
| <u>n =</u> | 126 | 126 | 126 | 1005 |

* p < .05; ** p < .01

An overall correlation (Pearson's *r*) of 0.28 (*p* < 0.000; 1005 ratings) was found between ORS and SRS total scores. Table 2 shows the correlations found at each of ten administrations of the instruments. All correlations are significant and of moderate magnitude, except for the low correlation at the first administration of the rating scales. Although ORS and SRS scores were moderately correlated, clients clearly discriminated between their views of the alliance (SRS) and their evaluations of daily functioning (ORS).

Reliability of the ORS and SRS

Table 3: Reliability (Cronbach's α) ORS and SRS

| Administration | <u>1st</u> | <u>2nd</u> | <u>3rd</u> | <u>1st to 10th</u> |
|---------------------------------|-----------------------|-----------------------|-----------------------|--|
| ORS | .84 | .91 | .88 | .91 |
| SRS | .86 | .89 | .90 | .90 |
| (s_{diff}) ORS = | 4.64 | 3.78 | 4.41 | 3.95 |
| 1.96(s_{diff}) ORS = | 9.09 | 7.40 | 8.64 | 7.73 |
| (s_{diff}) SRS = | 3.02 | 2.91 | 2.64 | 2.64 |
| 1.96(s_{diff}) SRS = | 5.91 | 5.70 | 5.17 | 5.17 |
| <u>n =</u> | 126 | 126 | 126 | 1005 |

Internal consistency. The homogeneity of the ORS and SRS was calculated using Cronbach's alpha. As Table 3 shows, internal consistency reliability (homogeneity) of both instruments is quite high for such brief instruments—alpha coefficients exceed 0.80. Table 3 also presents standard error of differences (s_{diff}) values. The standard error of difference formula estimates the dispersion of difference scores that would be expected by 'chance'; i.e., under the assumption that there is no actual difference between any two scores of the same client on different administrations of the same instrument (Jacobson & Truax, 1991; Hafkenscheid, 1994). If a change score exceeds the standard error of differences, multiplied by the z-value of the significance level sought (generally, $z = \pm 1.96$, two-sided, $\alpha = 0.05$; i.e., $1.96(s_{\text{diff}})$), this change score can be considered statistically reliable. Following Miller and Duncan (2004), the obtained internal consistency values were entered as reliability estimates in the standard error of differences formula. Calculated in this way, a score change of 7.73 can be considered statistically reliable for the Dutch ORS.

Comparison with American results. Miller et al. (2003) presented coefficient alpha values ranging from 0.87 to 0.93 for the American ORS (p. 95), whereas Duncan et al. (2003) reported a coefficient alpha of 0.88 (p. 8) for the American SRS. By and large, these values are within the same range as those found for the Dutch instruments. For the American ORS, a score difference of at least of 5 points (Miller & Duncan, 2004, p. 13) was obtained as an estimate of statistically reliable change.

Stability of ORS and SRS scores across treatment sessions. Using the same procedure as pursued in the American studies, stability of both ORS and SRS scores across treatment sessions were estimated by correlating scores from subsequent administrations of the instrument. Table 4 presents correlations between subsequent scores for the first four administrations of the instruments.

Table 4: Stability ORS and SRS scores across sessions

| Administration | <u>1st to 2nd</u> | <u>2nd to 3rd</u> | <u>3rd to 4th</u> |
|----------------|---|---|---|
| ORS | .54** | .63** | .16 |
| SRS | .49** | .65** | .16 |
| n = | 126 | 126 | 114 |

Statistically significant correlations (Pearson's r) between 0.49 and 0.65 are found for the instruments, when correlating subsequent scores obtained at the initial, second and third administration of the scales. In contrast, correlations between the third and fourth administration of the scales are low and non-significant.

Comparison with American results. For the ORS, Miller et al. (2003; p. 95, Table 3) reported what they call 'test-retest reliability' coefficients (Pearson's r) ranging from 0.49 to 0.66. An overall 'test-retest reliability' of $r = 0.64$ was reported for the SRS by Duncan et al. (2003; ibid, p. 9). Globally, the Dutch stability coefficients are within the same range as those obtained in the American psychometric studies, except for the low correlations between the third and fourth administrations of the Dutch ORS and SRS. It should be noted that correlations between subsequent administrations are an inappropriate (and too stringent) operational definition of test-retest reliability for instruments designed to be sensitive to client's perception of subjective change. Both state measures will tend show a pattern of declining correlations the first administration and each subsequent administration; this is referred to as autocorrelation.

Predictive Validity: ORS and SRS Scores over the Course of Treatment

The mean scores for both instruments turned out to be fairly stable across the first three administrations of the rating scales (see Table 1). In the worst scenario this could mean that no progress (in the alliance and/or in daily functioning) was achieved in this client group. A subgroup of 68 clients completed ten administrations of the ORS and SRS. Linear and curvilinear regression analyses were conducted to test treatment progress in this subgroup. SRS scores did not increase significantly across the ten administrations of the instrument (linear: $F = 0.39$, $p = 0.53$; logarithmic: $F = 0.11$, $p = 0.74$, $df = 1, 678$). In contrast, a statistically significant trend towards higher ORS scores over the course of treatment was established (linear: $F = 9.3$, $p = 0.002$; logarithmic: $F = 7.6$, $p = 0.006$, $df = 1, 678$).

In the same subgroup of 68 clients, the SRS was used as a predictor variable by consecutively correlating (using Pearson's r) SRS scores at each of the first three administrations of this alliance

measure with the tenth ORS administration as the outcome measure (criterion variable). SRS scores were not significantly correlated at the first and second administration (0.19 and 0.21) but the SRS did significantly correlate at the third administration (0.42; $p < 0.01$) with ORS ratings at the tenth administration. Thus, client's ratings of the alliance at the third administration of the SRS appear to be predictive of outcome in terms of future ORS scores, seven sessions ahead.

Absolute initial, second or third SRS scores did not predict the extent of ORS score changes from the first (pre-test) to the tenth (post-test) session ($-0.09 \leq r \leq 0.11$). Finally, a small, nearly significant ($p = 0.06$) positive correlation ($r = 0.23$) was found between the magnitude of SRS score changes from the first (SRS pre-test) to the third (SRS post-test) administration with the amount of ORS score changes from the first (ORS pre-test) to the tenth (ORS post-test) session.

Convergent validity: correlations of the ORS and SRS with therapist satisfaction (TSS)

An inspection of Table 5 reveals correlations (Pearson's r) between clients' ratings of the alliance (SRS) and their therapists' satisfaction with treatment sessions (TSS). Also, clients' ratings of their daily life functioning (ORS) were correlated with therapists' satisfaction (TSS). The high internal consistency reliability (Cronbach's α) of the TSS ($0.79 \leq \alpha \leq 0.85$; see Table 5) justified these analyses. With therapists forgetting to complete the TSS after treatment sessions, missing data amounted to 105 in these analyses, leaving 900 administrations in the aggregated analyses.

Table 5: Convergence TSS scores with ORS and SRS scores

| Administration | <u>1st</u> | <u>2nd</u> | <u>3rd</u> | <u>1st to 10th</u> |
|--------------------------|-----------------------|-----------------------|-----------------------|--|
| $r_{\text{ORS TSS}}$ | .11 | .14 | .22* | .22** |
| $r_{\text{SRS TSS}}$ | .27* | .21* | .07 | .16** |
| TSS: Cronbach's α | .79 | .84 | .85 | .83 |
| <u>n =</u> | 116 | 111 | 113 | 900 |

* $p < .05$; ** $p < .01$

Low but statistically significant positive correlations (Pearson's r) between clients' self-ratings and therapists' ratings are obtained at the third administration of the ORS and the first two administrations of the SRS. Overall, statistically significant positive correlations are found between ORS and TSS scores as well as between SRS and TSS scores. In the aggregated data set, the correlation between ORS and TSS scores is slightly higher than the correlation between SRS and TSS scores.

Convergence between client and therapist ratings is likely to be dependent on individual client-therapist pairs: in some pairs convergence may be strong, in other pairs convergence may be absent, and in some pairs the client and the therapist may even disagree in their evaluations of the alliance or treatment progress. Therefore, client (ORS and SRS) and therapist (TSS) rating were also correlated for each separate client-therapist pair. This analysis was limited to the 68 clients receiving ten administrations of the scales. Three pairs were lost in this analysis, because therapists did not complete the TSS, leaving 65 pairs.

Correlations between clients' scores on both instruments and TSS scores did indeed widely vary between pairs: from $r = -0.56$ to $r = 0.82$ for SRS-TSS correlations (Median $r = 0.14$); from $r = -0.74$ to $r = 0.80$ for ORS-TSS correlations (Median $r = 0.22$). At least moderate convergence ($r \geq$

0.40) occurred in 16 client-therapist dyads (24.6 %) for the SRS and in 19 dyads (29.2 %) for the ORS. At least moderate divergence ($r \leq -0.40$) was found in 5 client-therapist dyads (7.6 %) for both the SRS and the ORS.

Finally, average (ORS, SRS, and TSS) scores per client-therapist pair were correlated across all 65 pairs. The correlation between average SRS scores and average TSS scores was $r = -0.02$ (not significant) and the correlation between average ORS and TSS scores was $r = 0.27$ ($p = 0.03$). Again, convergence between ORS and TSS scores is higher than between SRS and TSS scores.

Discussion

All the problems typically associated with brief self-report tests (Boulet & Boss, 1991) apply to the ORS and its translations; for example, interpretation relies on clients' accurate assessment of their levels of distress and, on the SRS, 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. Clearly, evaluation of outcome via the ORS and the alliance with the SRS is far from comprehensive and does not contain multiple perspectives (e.g., therapists, outside judges, community criteria, etc.). Other limitations of the current study include the predominance of one therapist's clients in the sample and the possibility that the selected sample may not generalize to other Dutch mental health settings. Further research is needed to address these potential shortcomings.

The substantial benefits of feedback, supported by a growing empirical base including different measurement systems with varied populations in diverse contexts, suggests that the time has arrived for routine monitoring of outcome and the use of continuous feedback to tailor and improve psychotherapy services (Anker et al., 2009; Reese et al., 2009). Thanks to its high feasibility in clinical settings, the PCOMS approach of monitoring outcome via the ORS and the alliance with the SRS enjoys considerable popularity in the US. PCOMS is also rapidly disseminating to other parts of the world (e.g., Hafkenscheid, 2008, 2009b). Until now, psychometric data on non-English versions of the ORS and SRS have not been available.

The present study investigated the psychometric properties of the Dutch translations of the ORS and SRS, and examined cross-cultural differences with American studies. Results obtained with the Dutch versions were largely consistent with the American data. Applying the same American standards for the interpretation of psychometric findings, the internal consistency reliability and stability (test-retest) of the Dutch ORS and SRS were replicated.

Although the ORS and SRS scores were moderately correlated, clients clearly discriminated between their ratings of the alliance (SRS) and their evaluations of daily functioning (ORS). Some evidence was found for the predictive validity of SRS scores: SRS scores at the third administration significantly predicted ORS scores at the tenth administration. Unfortunately, no other predictive validity analyses correlating ORS (change) scores from SRS (change) scores reached statistical significance. It is important to note that not much change in general was evidenced with this clinical sample.

Normative data for the American and Dutch instruments diverged somewhat. Although the first session ORS scores were very similar in the American and Dutch samples, the later session scores were lower in the Dutch sample. The findings of differences with later ORS scores could reflect a general lack of change with this sample or it could represent cross-cultural differences. SRS scores in this Dutch sample were lower than those reported in the American studies, suggesting that cross-cultural differences may exist between American and non-American respondents. These scoring differences may also reflect sample differences. Other investigations in both American and abroad are finding similar differences to the original SRS normative data. An investigation of a Norwegian sample, for example, found similar differences on the SRS (Anker, personal communication, May 14,

2009) as reported in the Dutch sample, suggesting that clients may score the SRS lower than the original sample.

Finally, the low correlations between self-ratings and therapist ratings are in line with findings that clients' and therapists' assessment of the quality of the therapeutic alliance may differ to a considerable extent (e.g., Horvath & Bedi, 2002). The results of the present study confirm that clients and therapists have their own, unique perspective on the alliance and progress. Although it is the client's evaluation of the alliance that predicts outcome, there may be value in supplementing the ORS and SRS with a simple instrument like the TSS to further enhance conversations between clients and therapists.

The results of this investigation of the Dutch ORS and SRS suggest that they function similarly to the American instruments. Replication, however, in different client samples and other non-English translations are sorely needed to further encourage the spread of a continuous feedback approach. Continuous feedback has been repeatedly shown to improve outcomes (Duncan, 2010; Lambert, 2010) by individualizing psychotherapy based on treatment response and client preference, and could alleviate the drop-out problem as well as treatment and therapist variability. Additional validation of these very feasible instruments can serve to bring more clinicians on board to the benefits of systematic incorporation of client feedback.

References

- Anker, M., Duncan, B.L. & Sparks, J.A. (2009). Using client feedback to improve couple outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, 77, 693-705.
- Boulet, J., & Boss, M. (1991). Reliability and validity of the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology*, 3, 433-437.
- Bringhurst, 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.
- de Jong, K., Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, P., & Heiser, W.J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy*, 14, 288-301.
- Duncan, B. (2010). *On becoming a better therapist*. Washington, DC: American Psychological Association.
- Duncan, B.L., Miller, S.D., Sparks, J.A. (2004). *The heroic client: a revolutionary way to improve effectiveness through client-directed outcome informed therapy*. San Francisco: Jossey-Bass.
- Duncan, B.L., Miller, S.D., Sparks, J.A., Claud, D.A., Reynolds, L.R., Brown, J. & Johnson, L.D. (2003). The Session Rating Scale: psychometric properties of a "working" alliance measure. *Journal of Brief Therapy*, 3, 3-12.
- Duncan, B.L., Miller, S.D., Wampold, B.E. & Hubble, M.A. (2010). *The heart and soul of change: Delivering what works (2nd Ed.)*. Washington DC: American Psychological Association.
- Duncan, B., & Sparks, J. (2010). *Heroic clients, heroic agencies: Partners for change (2nd Ed.)*. Ft. Lauderdale: HSCP Press, www.heartandsoulofchange.com.
- Duncan, B., Sparks, J., Miller, S., Bohanske, R., & Claud, D. (2006). Giving youth a voice: A preliminary study of the reliability and validity of a brief outcome measure for children, adolescents, and caretakers. *Journal of Brief Therapy*, 5(2), 66-82.
- Evans, C., Connell, J., Barkham, M., Mellor-Clark, J., McGrath, G. & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180, 51-60.

- Hafkenscheid, A. (1994). *Rating scales in treatment efficacy studies: individualized and normative use*. Unpublished dissertation. Groningen: State University of Groningen, the Netherlands.
- Hafkenscheid, A. (2008). *Routine Process Monitoring*: ervaringen uit de praktijk. *Tijdschrift Cliëntgerichte Psychotherapie*, 46, 327-345.
- Hafkenscheid, A. (2009a). The impact of psychotherapy sessions: internal structure of the Dutch Session Evaluation Questionnaire (SEQ). *Psychology and Psychotherapy: Theory, Research and Practice*, 82, 99-111.
- Hafkenscheid, A. (2009b) *Routine Process Monitoring* (RPM) in partnerrelatiebehandelingen. *Directieve Therapie*, 29, 5-25.
- Hannan, C., Lambert, M.J., Harmon, C., Nielsen, S.L., Smart, D.M., Shimokawa, K. & Sutton, S.W. (2005). A lab test and algorithms for identifying patients at risk for treatment failure. *Journal of Clinical Psychology: In Session*, 61, 155-163.
- Horvath, A.O & Bedi, R.P. (2002). The alliance. In: J.C. Norcross (Ed.). *Psychotherapy relationships that work: therapist contributions and responsiveness to patients* (pp. 37-69). New York: Oxford University Press.
- Jacobson, N.S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Lambert, M. J. (2010). "Yes, it is time for clinicians to routinely monitor treatment outcome." In B.L. Duncan, S.D. Miller, B.E. Wampold, & M.A. Hubble (Eds.) *The heart and soul of change. Delivering what works* (pp. 239-268). Washington, DC: APA Press
- Lambert, M.J., Burlingame, G.M., Umphress, V.J., Hansen, N.B., Vermeersch, D., Clouse, G. & Yanchar, S. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 249-258.
- Lambert, M.J. & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In: M.J. Lambert (Ed). *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 139-193). New York: Wiley.
- Miller, S.D. & Duncan, B.L. (2004). *The Outcome and Session Rating Scales: Administration and scoring manual*. Chicago, IL: Institute for the Study of Therapeutic Change.
- Miller, S.D., Duncan, B.L., Brown, J., Sparks, J.A. & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity and feasibility of a brief visual analogue measure, *Journal of Brief Therapy*, 2, 91-100.
- Miller, S.D., Duncan, B.L., Sorrell, R. & Brown, G.S. (2005). The Partners for Change Outcome Management System. *Journal of Clinical Psychology*, 61, 199-208.
- Reese, R.J., Norsworthy, L., & Rowlands, S. (2009). Does a continuous feedback model improve psychotherapy outcomes? *Psychotherapy*, 46, 418-431.
- Samstag, L.W., Batchelder, S.T., Muran, J. C., Safran, J.D., & Winston A. (1998). Early identification of treatment failures in short-term psychotherapy: an assessment of therapeutic alliance and interpersonal behaviour. *Journal of Psychotherapy Practice and Research*, 7, 126-143.
- Stiles, W.B., Reynolds, S., Hardy, G.E., Rees, A., Barkham, M. & Shapiro, D.A. (1994). Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *Journal of Counseling Psychology*, 41, 175-185.
- Tracey, T.J. (1992). *Client Satisfaction Scale (CSS) and Therapist Satisfaction Scale (TSS) manual*. Tempe: Arizona State University.