The Influence of Demand Characteristics and Social Desirability on Clients' Ratings of the Therapeutic Alliance

Robert J. Reese,[1] J. Arthur Gillaspy Jr.,[2] Jesse J. Owen,[3] Kevin L. Flora,[1] Linda C. Cunningham,[1] Danielle Archie,[2] and TroyMichael Marsden[2]

[1] University of Kentucky
[2] University of Central Arkansas
[3] University of Louisville

Objective: To examine demand characteristics, social desirability on clients' rating of working alliance using the Session Rating Scale (SRS; Miller, Duncan, & Johnson, 2000). **Method:** Clients ($N = 102$) at two sites were randomly assigned to one of three alliance feedback conditions: (a) IF—SRS completed in presence of therapist and the results discussed immediately afterward; (b) Next Session Feedback—SRS completed alone and results discussed next session; or (c) No Feedback—SRS completed alone and results not available to therapist. Clients completed the SRS for the first three sessions of treatment. **Results:** No statistically significant differences in SRS scores across the feedback conditions were found. Additionally, the analysis showed that SRS scores were not correlated with a measure of social desirability but were correlated with an established alliance measure. **Conclusions:** The results indicate that alliance scores were not inflated due to the presence of a therapist or knowing that the scores would be observed by the therapist. © 2013 Wiley Periodicals, Inc. J. Clin. Psychol. 69:696–709, 2013.

Keywords: client feedback; therapeutic alliance; therapeutic relationship; patient-focused research

Continuously monitoring client response to treatment using standardized measures, a process termed *client feedback*, has become a recommended practice in the delivery of psychotherapy treatment (Ackerman et al., 2001; American Psychological Association [APA] Presidential Task Force on Evidence Based Practice, 2006; Kazdin, 2007; Lambert, 2010). Over a decade's worth of research has consistently supported the efficacy of client feedback to improve treatment outcomes with a variety of clients across a range of treatment settings (e.g., Anker, Duncan, & Sparks, 2009; Reese, Norsworthy, & Rowlands, 2009; Shimokawa, Lambert & Smart, 2010).

There are several formal systems to collect client feedback (e.g., Crits-Christoph et al., 2010; Knobloch-Fetters, Pinsof, & Mann, 2007; Lambert et al., 2004; Miller & Duncan, 2004;). At the core of these systems is the session by session tracking of client distress/well-being, which provides the therapist a baramoter for how clients are progressing in therapy (or not progressing). For instance, if the client indicates progress (i.e., symptom distress is diminishing), therapists receive affirmation that their approach has been adaptive for the client. More importantly, if the client is not progressing, therapists can modify treatment or address issues within the therapeutic relationship before clients deteriorate or dropout (Duncan, 2010; Lambert et al., 2004).

Two client feedback systems that have solid research support are the Outcome Questionnaire Quality Management System (OQ System; Lambert et al., 2004) and the Partners for Change Outcome Management System (PCOMS; Miller & Duncan, 2004). In both systems clients complete brief outcome measures (Outcome Questionnaire 45 or Outcome Rating Scale, respectively) via paper-pencil, computer, or handheld device before each session. The results are immediately transmitted to their therapist along with treatment suggestions, depending on the client's progress (improving, no change, or deterioration). Additionally, in some client feedback systems, there are opportunities for clients to provide feedback about the therapuetic alliance (e.g., Duncan, Miller, & Sparks, 2004; Knobloch-Fedders, Pinsof, & Mann, 2007; Whipple

et al., 2003). Similar to tracking client progress, the alliance is typically assessed at the end of each session and results are discussed with clients.

Support for client feedback has been promising. For instance, Shimokawa et al. (2010) conducted a meta-analysis of six randomized clinical trials (RCT) comparing the OQ System to treatment as usual (TAU; Harmon et al., 2007; Hawkins et al., 2004; Lambert et al., 2001, 2002; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Whipple et al., 2003). These studies were further subdivided by type of feedback condition: feedback provided to therapist only, feedback provided to therapist and client, feedback that included the clinical support tools (CSTs) such as the alliance, stage of change. For clients identified as not-on-track or at-risk for treatment failure, the average effect size when feedback was provided to therapists only and to therapists and clients was $r = .25$. When the CSTs were included in the feedback condition, the average effect size was $r = .33$. More recently, Lambert and Shimokawa (2011) analyzed three RCTs of PCOMS (Anker et al., 2009; Reese et al., 2009; Reese Toland, Slone, & Norsworthy, 2010). These studies included individual and couples clients at a community mental health center, a university counseling center, and a training clinic, respectively. All clients (those at risk for treatment failure and those who were progressing normally) experienced significantly improved client retention and outcome versus TAU (effect size for outcome, $r = .23$).

Notwithstanding these promising strides for client feedback, there have been questions about the validity of clients' ratings of the alliance when provided in the presence of the therapist or when the therapist will have access to clients' scores. The focus of the current study is the Session Rating Scale (SRS; Miller et al., 2000), the alliance feedback component of the PCOMS. The SRS is an ultra-brief measure of the alliance (four items), which is consistent with a general trend in the field to utilize shorter measures of the alliance (e.g., Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011; Hatcher & Gillaspy, 2006; Tracey & Kokotovic, 1989). One advantage of the brevity of the SRS is increased feasibility and usage. In a study conducted at two clinics, Duncan et al. (2003) found the utilization rate of the SRS was 96% compared with 29% for the 12-item Working Alliance Inventory-Short (WAI-S; Tracey & Kokotovic, 1989).

The SRS was intentionally designed to provide immediate feedback to therapists about the alliance as experienced in each session. Clients provide this feedback and discuss their ratings face-to-face with therapists, allowing therapists to monitor the alliance in realtime for any disconnects or ruptures and immediately intervene. Likewise, if clients rate the alliance as strong, therapists learn valuable information about aspects of the alliance that are working for the client.

The utility of the SRS, however, depends on the reliability and validity of its scores. As noted by Crits-Christoph et al. (2011)"

> It would obviously be useful to measure the alliance repeatedly. To the extent that this presents a burden on patients, particularly if an outcome instrument is already administered at treatment sessions, very brief measures of the alliance might have greater clinical utility, assuming the reliability and validty of such measures are adequate. (p. 275)

Therefore, the psychometric properties of SRS scores, especially given it's brevity, are of concern. In a recent review of PCOMS studies, Gillaspy and Murphy (2011) reported the average internal consistency of SRS scores across five studies equaled .92 (range .88 to .96). SRS scores also exhibit moderate evidence of concurrent validity with longer alliance measures; $r = .48$ with the Helping Alliance Questionnaire-II (HAQ-II; Luborsky et al., 1996) and $r = .58$ with the WAI-S (Tracey & Kokotovic, 1989). The SRS appears to have some overlap with the other alliance measures but may also tap a different alliance conceptualization. The predictive validity of SRS scores has been supported by two studies. Duncan et al. (2003) found a correlation of $r = .29$ between early SRS scores and treatment outcome, which is consistent with previous research alliance-outcome research (Horvath, Del Re, Flückiger, & Symonds, 2011). More recently, Anker, Owen, Duncan, and Sparks (2010) reported third session SRS scores predicted outcome beyond early symptom change ($d = 0.25$).

Although the initial psychometrics properties for the SRS are encouraging, one characteristic of SRS scores that has not been examined is the presence of social desirability effects. Social desirability refers to "the tendency for people to distort their responses in a manner that portrays them in a positive light" (Leary, 2001, p. 409). In the context of the alliance, this may be especially relevant given the potential demand characteristics with clients being asked to rate the quality of their relationship with their therapist present; clients may not want to respond in a negative manner. Although social desirability is frequently an issue in psychological measurement, Tryon, Blackwell, and Hammel (2007) noted that it is seldom addressed in alliance assessment. One reason for this is that client responses on alliance measures are rarely seen or used by clinicians in research studies. Clients are typically assured that their answers will not be shared with their counselor, thus minimizing concern about social desirability.

The concern that clients who complete the SRS face-to-face with their therapists may not be candid in their appraisal of the quality of the relationship has been acknowledged. Duncan (personal communication, August 5, 2011) and Murphy (personal communication, September 1, 2010) report that questions about the honesty of clients' responses on the SRS are frequently voiced at PCOMS training workshops. In addition, Reese et al. (2010) cautioned that a social desirability response set on the SRS may be a serious limitation of feedback research using the PCOMS. Specifically, one requirement for an effective feedback system is that the feedback source be credible (Luborsky et al., 1996). Thus, if using the SRS to monitor client feedback about the alliance is to become part of routine therapy practice, then it is imperative to understand the role of demand characteristics and social desirability on the validity of SRS scores.

We had four research questions for the current study. First, does altering the feedback process result in differences in SRS scores? This question was evaluated by randomly assigning clients to one of three treatment conditions: (a) Immediate Feedback–clients completed and immediately discussed the SRS scores in the presence of their therapist; (b) Next Session Feedback–clients completed the SRS alone and discussed the SRS scores the next session; or (c) No Feedback–clients completed the SRS alone and their therapist did not see the scores. Second, does altering the feedback process result in different rates of change in SRS alliance scores across the first three sessions of therapy? Directional hypotheses were not proposed because there is no foundational evidence for supporting a specific supposition. Third, we tested whether clients' ratings of social desirability were associated with their alliance scores. Fourth, we investigated concurrent validity for SRS scores with a more established alliance measure, the Working Alliance Inventory-Short Revised (WAI-SR; Hatcher & Gillaspy, 2006). An established measure was selected for comparison because of continued psychometric concern regarding the SRS's brevity, specifically with regard to concurrent validity (Gillaspy & Murphy, 2011) and the SRS being utilized in a different manner than prescribed in the PCOMS manual (Miller & Duncan, 2004).

Method

Participants

Clients. A total of 102 clients were recruited from a small, private southeastern university counseling center (CC1; $n = 22$) and a counseling center at a medium-to-large, state southeastern university (CC2; $n = 80$). Initially, 137 clients consented to participate in the study during an intake session (no one declined to participate), but 35 clients did not return for treatment. The final client sample comprised undergraduate students who were primarily female (79.4%) and White (75.5%; 11.8% Hispanic/Latino; 4.9% African American; 2.9% American Indian/Alaskan Native; 1.0% Asian American, 1.0% indicated "Other," and 2.9% did not indicate ethnicity/race). The mean age for clients was 21.92 (standard deviation [SD] = 5.29) with a range of 18 to 51. On a measure of general distress, the Outcome Rating Scale (ORS; Miller & Duncan, 2000), both sites had mean ORS scores considered in the clinical range (CC1 = 22.30 [SD = 6.81] and CC2 = 24.03 [SD = 6.03]). The difference between the scores at each site was not statistically significant.

Therapists. Therapy was provided by 17 therapists (CC1 = 5; CC2 = 11). The therapists at both counseling centers were either practicum students (CC1 = 3; CC2 = 6) or professional, licensed staff. The majority of the therapists were female ($n = 12$) and White (56.25%; African American = 18.75%; Asian/Indian = 25.0%). Therapists saw a range of 1–17 clients (Median = 4; Mode = 3). Theoretical approaches for therapists varied, including cognitive-behavioral, solution-focused, psychodynamic, narrative, and emotion-focused therapy.

Measures

Marlowe-Crowne Social Desirability Scale–Short Form (SDS; Ballard, 1992). The SDS is a 13-item abbreviated version of the 33-item Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). The measure uses a true-false format that is designed to identify those who are answering items in a manner that portrays themselves in a favorable or socially desirable manner. The SDS is designed to measures a stable trait-like behavior wherein individuals minimize weakness and exaggerate strengths when completing attitudinal and personality-based assessments. Item examples include, "I sometimes feel resentful when I don't get my own way" or "I'm always willing to admit it when I make a mistake." The SDS is scored by summing the 13 items (keyed 0 or 1), with higher scores indicating more social desirability. Reliability estimates for the SDS have been modest, ranging from .62 (Loo & Thorpe, 2000) to .66 (Loo & Loewen, 2004). Although reliability estimates are lower than the full-scale version, the SDS has evidence of stronger construct validity than the full version (Ballard, 1992; Loo & Thorpe, 2000). The internal consistency for the current sample was .64.

ORS (Miller & Duncan, 2000). The ORS is a four-item, self-report measure of psychological distress designed to evaluate session-to-session progress in therapy. Clients rate their level of distress on items adapted from the three areas of the OQ45 (Lambert et al., 1996). Specifically, clients respond on a visual analog scale to how they are doing *Individually* (personal well-being), *Socially* (work, school, friendships), *Interpersonally* (family, close relationships), and *Overall* (general sense of well-being). Clients make a mark on each of the four analog scales that are 10 cm in length, with marks near the left end of the scale indicating more distress and marks near the right end of the scale indicating less distress. A ruler or template is then used to measure the distance from the left end of the scale to the client's mark. The score is recorded for each item to the nearest millimeter and are then totaled, ranging from 0 to 40. Lower scores reflect more distress.

The internal consistency estimated with the ORS for the current sample was .77, .88, and .90 for the first three sessions, respectively. Anker et al. (2009) and Reese et al. (2009) found reliability estimates comparable with the last two sessions. Evidence of concurrent validity for scores derived from the ORS is based on Pearson correlations with scores on other established outcome measures, including the Clinical Outcomes in Routine Evaluation (Barkham et al., 2001; $r = .67$; Miller & Duncan, 2004) and the OQ45 ($r = .59$; Miller, Duncan, Brown, Sparks, & Claud, 2003).

SRS (Miller et al., 2000). The SRS comprises four items designed to measure the therapeutic alliance using a visual analog scale. The SRS was developed based on alliance theory and research (Bordin, 1979; Gaston, 1990; Hatcher & Barends, 1996). The first three items are based on Bordin's (1979) conceptualization of the working alliance, assessing the therapeutic relationship ("I felt heard, understood, and respected"), the goals and topics covered in therapy ("We worked on and talked about what I wanted to work on and talk about"), and the method or approach used in therapy ("The therapist's approach is a good fit for me"). The last item is based on Hatcher and Barend's (1996) confident collaboration dimension of the alliance and evaluates the overall rating of the session ("Overall, today's session was right for me"). Clients make a mark on each of the four analog scales that are 10 cm in length, with marks near the left end of the scale indicating less satisfaction and marks near the right end of the scale indicating higher satisfaction for each item. The SRS scoring process is identical to the ORS with the individual items scored and totaled (0 to 40). Lower scores reflect less satisfaction. The PCOMS

manual recommends that any total score below 36 or any individual item rated below 9 should be discussed to identify ways to improve upon the score. This cut score was derived based on over 15,000 administrations that found roughly only a quarter of respondents reported a score below 36 (Miller & Duncan, 2004).

For the current sample, coefficient alpha was estimated at .91, .91, and .92 for the first three sessions of the SRS, respectively.

WAI-SR (Hatcher & Gillaspy, 2006). The WAI-SR is a 12-item measure of working alliance as conceptualized by Bordin (1979). The WAI-SR was constructed using factor analysis and item response theory methods to maximize the differentiation between the Bond, Task, and Goal dimensions of working alliance. Clients respond to items on a 5-point Likert scale. Items are positively worded, however the valence of half the responses are reversed; 1 (*always*) to 5 (*seldom*). Example items include, "[*The therapist*] and I respect each other" (Bond), "As a result of these sessions I am clearer as to how I might be able to change" (Task), and [*The therapist*] and I collaborate on setting goals for my therapy" (Goal). The WAI-SR is a fairly widely used working alliance instrument with demonstrated score reliability and validity (Hatcher & Gillaspy, 2006; Munder, Wilmers, Leonhart, Linster, & Barth, 2010). For the current sample, Cronbach's alpha of WAI-SR scores for the first three sessions equaled .93, .92, .92, respectively.

Procedure

All participants were new clients. Clients were randomly assigned before their first session to one of three conditions where the SRS was completed at the end of the first three sessions: (a) Immediate Feedback (IF)–in the presence of their therapist and the results were discussed immediately afterward (as prescribed by the PCOMS protocol); (b) Next Session (NS) Feedback—after their therapist left the room and the results were discussed the next session; or (c) No Feedback (NF)—after the therapist left the room and the results not made available for therapist review or discussion. Clients also completed the WAI-SR at the end of their first three sessions and the SDS after the first session (therapists did not have access to these completed measures). The WAI-SR was administered after the SRS for all three sessions, and the SDS after the WAI-SR for the first session. All measures were administered using a paper-pencil format. A randomized block design was used to help control for therapist effects, and all therapists were trained to implement the PCOMS. Training at each site consisted of a minimum of $1^{1}/_{2}$ hours of formal instruction in the rationale for utilizing client feedback and the mechanics of administering, scoring, and interpreting the ORS and SRS. The same study protocol was used at both sites.

Clients consented to the study at an intake session before beginning treatment and were told that they would be completing four measures to facilitate understanding of the therapy process; clients were not given information about specific hypotheses. At the first session, therapists provided explicit instructions on when the measures would be completed and if the therapist would see the completed measures at the first session. This included providing a brief rationale for using the ORS and SRS to track treatment progress. For all treatment conditions, therapists used the ORS at the beginning of each session as consistent with the PCOMS manual. After clients completed the ORS (approximately 1 minute), therapists scored the items in the session. The total score was charted on a graph that indicated each individual client's progress across treatment. Therapists utilized the ORS as consistent with the PCOMS manual. This was done with all treatment conditions. At the end of each session (approximately 3–4 minutes left in the session), clients completed the SRS (approximately 1 minute) as indicated for each treatment condition described earlier.

Depending on the treatment condition, participants were given explicit direction on how the SRS would be utilized. In the IF condition, the client completed the SRS with the therapist present at the end of the session. The therapist then scored and discussed the SRS results with the client. Scores below 36 or an individual item rated less than nine were discussed per the PCOMS manual. In the NF condition, to emphasize anonymity, clients were given an envelope for the SRS, which they sealed and placed in either a drop-box (CC1 and CC2 sites) outside of the therapy room or the mailbox of the first author (TC site). Clients in the NS condition

also completed the SRS and placed the measure in either the drop-box or mailbox. Scores were entered into the research database and then returned to the therapist for use the next session. After completing the SRS, regardless of treatment condition, the WAI-SR (at the end of each of the first three sessions) and SDS (at the end of the first session only) were completed by clients and then placed in a sealed envelope and also placed by the client in the aforementioned drop-box or mailbox. Clients were reminded that these measures would not be seen by the therapist.

Data Analysis Plan

By design, the current study has dependent data with time (session 1 to session 3) nested within clients who were nested within therapists (multiple clients per therapist). As such, we conducted our main analyses with three-level multilevel models (MLM) to account for the interdependencies in data. Psychotherapy studies typically have nested data structures (e.g., clients nested with therapists, repeated sessions nested within clients who are then nested within therapists). The concern with nested data structures is that the assumption of independence of observations is violated, which is necessary for traditional techniques like analysis of variance (Peugh, 2010). Disregarding this concern may result in biased parameter estimates (i.e., means, variances, and covariances) and increased probability of Type I errors.

In this study, we were mainly interested in whether Feedback condition would affect the mean SRS scores (i.e., the intercept). Thus, the main model included SRS scores as the outcome variable and Time as the only level 1 predictor. At level 2, we included Feedback conditions and Social Desirability scores (note the SDS was only assessed once prior to therapy). There were no therapist level variables in the model. The full final model was:

$$Y_{tij} = \gamma_{000} + \gamma_{010}(SDS) + \gamma_{020}(IF - Feedback) + \gamma_{030}(NS - Feedback) + \gamma_{100}(Time) + [error] \quad (1)$$

Where Y_{tij} is the SRS score at time point three (i.e., session 3) for client i who was treated by therapist j, SDS is the social desirability score for clients, IF-Feedback is the group of clients in the IF condition, NS-Feedback is the group of clients in the NS Feedback condition (note in this model the NF condition was the comparison group), and Time was coded –2 for session 1, –1 for session 2, and 0 for session 3. The error terms were bracketed here for brevity.

Next, we tested whether Feedback conditions would affect the degree of change in SRS scores. As such, we replicated Equation 1 but included the same predictors at Level 2 in the prediction of Time at Level 1. This analysis is typically described as "slopes as outcomes" model (Raudenbush & Bryk, 2002). Specifically, the full final model was:

$$\begin{aligned} Y_{tij} = {} & \gamma_{000} + \gamma_{010}(SDS) + \gamma_{020}(IF - Feedback) + \gamma_{030}(NS - Feedback) \\ & + \gamma_{100}(Time) + \gamma_{110}(SDS * Time) + \gamma_{120}(IF - Feedback * Time) + \gamma_{130} \\ & (NS - Feedback * Time) + [error] \end{aligned} \quad (2)$$

In this model, the Variable*Time (e.g., SDS*Time) represents the cross-level interaction, predicting the slope or degree of change in SRS scores by each variable.

Regarding the random effects, we initially tested whether Time (or the slope of SRS scores) would vary across clients and therapists. The slope was significant for clients ($p < .05$), but not therapists ($p > .05$). We also attempted to allow Feedback conditions to vary across therapists; however, the model did not converge, which is likely due to primarily having only one client per condition per therapist. As such, we fixed all of the effects across therapists, but allowed the slope for Time to vary across clients. Multilevel models were conducted using hierarchical linear modeling version 6 (Raudenbush, Bryk, Cheong, & Congdon, 2005).

Table 1

Means and Standard Deviations for Clients' SRS, WAI-SR, and ORS Scores for Sessions 1 to 3

	No Feedback			Feedback Next Session			Feedback in Session		
	S 1	S 2	S 3	S 1	S 2	S 3	S 1	S 2	S 3
SRS	36.18	37.34	36.79	33.91	36.39	37.10	36.00	36.65	37.28
	(4.58)	(2.98)	(2.91)	(6.40)	(3.64)	(3.48)	(3.43)	(3.49)	(5.15)
WAI-SR	51.11	52.63	52.67	49.46	53.69	55.89	51.07	52.83	56.33
	(7.55)	(6.47)	(7.38)	(7.22)	(5.57)	(5.35)	(8.68)	(7.95)	(4.67)
ORS	22.62	28.64	29.78	22.75	28.16	31.60	25.12	29.56	33.02
	(6.60)	(5.90)	(5.73)	(6.04)	(5.86)	(5.74)	(7.63)	(6.72)	(6.28)
SDS	6.06	–	–	4.90	–	–	5.97	–	–
	(2.65)			(2.55)			(2.68)		
Ns	27–33	19–27	12–20	18–29	13–22	9–17	27–40	18–29	15–25

Note. S1 to S3 = Session 1, 2, and 3; SRS = Session Rating Scale; WAI-SR = Working Alliance Inventory-Short Revised; ORS = Outcome Rating Scale; SDS = Social Desirability Scale.
Numbers outside of the parentheses reflect the mean scores and the numbers within the parentheses reflect the standard deviation scores.

Results

Table 1 provides an overview of the descriptive statistics for the SRS, WAI-SR, ORS, and SDS of the three treatment conditions (IF, NS Feedback, and NF). The range of participants at each session reflects completion of the measures. For example, at session 1 in the NF condition, six of the 33 total participants had a missing data point. As can be observed, the mean scores and standard deviations for the SRS and WAI-SR indicate most clients rated therapeutic alliance high. However, the percentage of SRS session scores ≤ 36 (32.89%) was actually higher than the normative sample used to derive the clinical cut score (approximately 24%; Miller & Duncan, 2004). To provide further context, the mean SRS score (session 1) was a little higher than the initial mean SRS score found with a Norwegian sample of individuals in couple therapy (33.48; Anker et al., 2009). For the ORS, the overall mean score was in the clinical range (< 25), which is consistent with previous research using clinical samples (Anker et al., 2009; Reese et al., 2009, 2010).

We evaluated each of the measures for normality using recommendations from West, Finch, and Curran (1995) for identifying skewness (>2) and kurtosis (>7). The SRS (session 3) met criteria for skewness (-3.31) and kurtosis (14.91), but none of the other measures across the three sessions were considered skewed or kurtotic. Therefore, no data transformations were conducted. The problem of having normally distributed alliance measures is consistent with most alliance research (Tryon et al., 2007). We also analyzed the SRS data for each session to investigate if missing data was random across treatment conditions that could have unduly influenced the findings. Reasons for missing data could have included the therapist forgetting to administer the measures or clients not attending a subsequent session. A chi-square analysis indicated that SRS missing data across treatment conditions was not statistically significantly different at session 1, χ^2 (degree of freedom [*df*] = 2, $N = 102$) = .21, $p > .05$; session 2, χ^2 (*df* = 2, $N = 102$) = .08, $p > .05$; or session 3, χ^2 (*df* = 2, $N = 102$) = 1.37, $p > .05$. MLM is robust to violations of missing data (Raudenbush & Bryk, 2002); therefore, no missing data replacement methods were used.

To evaluate the first research question–whether SRS scores would vary based on Feedback condition–we ran the model described above in Equation 1. As seen in Table 2, clients' SRS scores did not significantly vary as a function of their Feedback Condition assignment. Specifically, the difference between clients in the IF condition and NS Feedback condition versus the NF condition were small-sized effects, *ds* = 0.03, 0.15, respectively. The difference between the IF condition versus NS Feedback condition was also a small-sized effect, *d* = –0.02. Only the slope for SRS scores was statistically significant, suggesting that clients increased their SRS scores

Table 2
Summary of Fixed Effects for the Multilevel Model Predicting SRS Intercept Score

SRS	Coefficient (SE)	t-ratio	p-value	D
Intercept	36.83 (0.59)	62.57	< .001	–
SDS	0.12 (0.12)	1.02	.31	0.03
IF vs. NF	0.58 (0.70)	0.83	.41	0.03
NS vs. NF	0.68 (0.88)	0.77	.44	0.15
IF vs. NS	−0.10 (0.69)	−0.14	.89	−0.02
Slope	0.68 (0.24)	2.78	.013	0.15

Note. SRS = Session Rating Scale; SE = standard error; IF = Immediate Feedback Condition; NF = No Feedback Condition; NS = Next Session Feedback Condition; SDS = Social Desirability Scale. d = effect size, 0.20 = small effect, 0.50 = medium effect, and 0.80 = large effect. Effect sizes were calculated by dividing the coefficient from the SD (across the three sessions) for the SRS.

Table 3
Summary of Fixed Effects for the Multilevel Model Predicting SRS Slope Score

SRS	Coefficient (SE)	t-ratio	p-value	d
Slope	0.50 (0.46)	1.09	.28	0.11
SDS	0.09 (0.10)	0.89	.37	0.02
IF vs. NF	−0.21 (0.62)	−0.34	.73	−0.05
NS vs. NF	0.86 (0.66)	1.30	.20	0.19
IF vs. NS	−1.07 (0.64)	−1.68	.10	0.24

Note. SRS = Session Rating Scale; SE = standard error; IF = Immediate Feedback Condition; NF = No Feedback Condition; NS = Next Session Feedback Condition; SDS = Social Desirability Scale. d = effect size, 0.20 = small effect, 0.50 = medium effect, and 0.80 = large effect. Effect sizes were calculated by dividing the coefficient from the SD (across the three sessions) for the SRS.

0.68 points per session, which is notable that the scores did significantly change even though they started relatively high. Thus, there was no support indicating that feedback condition had an effect on clients' scores on the SRS. In regard to our third research question, clients' SDS scores were not significantly associated with SRS scores ($d = 0.03$). We also tested whether SDS would interact with the Feedback condition in the prediction of SRS scores and the rate of change in the SRS. There were no significant interaction effects for the intercept (SRS scores at third session) or the slope (rate of change in SRS scores) across the conditions ($ps > .10$).

Next, we tested whether the change in SRS scores over the three sessions would vary based on Feedback Conditions using Equation 2 listed above (research question 2). For brevity, Table 3 only provides a summary of the results for the cross-level interactions (e.g., Time x IF condition). Similar to above, there was no significant association between any of the predictor variables in the prediction of the SRS slope. That is, clients in the Feedback conditions did not significantly differ from one another in the changes in their SRS scores.

Although not shown in our table, we conducted the same models with WAI-SR as the dependent variable. The results were consistent, insofar that none of the conditions significantly varied from one another in the intercept or slope ($ps > .10$). However, similar to the models with the SRS, the slope for the WAI-SR was statistically significant ($\gamma = 2.13$, standard error [SE] = .60, $p < .001$, $d = 0.29$). These results are not as surprising as clients were told that their therapist would not have access to the WAI-SR scores and therapists did not discuss these scores with their clients.

We also evaluated the concurrent and discriminant validity of the SRS scores by computing a correlational analysis between the SRS and the SDS (discriminant validity) and the WAI-SR (concurrent validity). As can be observed in Table 4, low correlations were found between the SRS and the SDS ($r = .05$) at session 1 and moderate to strong correlations with the WAI-SR

Table 4
Bivariate Correlations for the SDS, WAI-SR, and SRS

Measure	1	2	3	4	5	6	7
1. SDS	-						
2. WAI-SR1	.01	-					
3. WAI-SR2	.01	.81***	-				
4. WAI-SR3	.02	.72***	.67***	-			
5. SRS1	.05	.65***	.44**	.42*	-		
6. SRS2	.17	.41**	.57***	.20	.41**	-	
7. SRS3	.02	.52**	.38*	.62***	.39**	.61***	-

Note. SDS = Social Desirability Scale; WAI-SR1 to WAI-SR3 = Working Alliance Inventory – Short Form (Sessions 1–3); SRS1 to SRS3 = Session Rating Scale (Sessions 1–3).
*$p < .05$. **$p < .01$. ***$p < .001$.

(ranging from $r = .57$ [session 2] to $r = .65$ [session 3]), providing evidence for both discriminant and concurrent validity.

Discussion

This study investigated the concern that clients inflate scores on therapeutic alliance measures because of demand characteristics or social desirability. Specifically, we had three major findings. The first finding was that clients' alliance scores were not different when the measures were completed in the presence of their therapist or knowing that the results would be discussed with their therapist. Another way of stating this is that anonymity did not lead to lower or higher scores. Such a finding should lessen concern with *how* alliance measures are administered.

A common concern of client feedback systems that utilize alliance measures (Miller & Duncan, 2004; Pinsof et al., 2007) is that clients are less likely to be honest in their appraisal of the therapeutic alliance when completing the measure in front of their therapist and then discussing the results (Duncan, personal communication, August 5, 2011). This process certainly heightens the potential for demand characteristics to influence scores. Would a client risk "hurting the feelings" of a therapist, and actually voice unhappiness with a session to the therapist's face? This study does not definitely answer this, but the results indicate that therapist presence and the subsequent discussion of the results did not lessen the possibility of expressing dissatisfaction.

There are few studies that have investigated this phenomenon directly, and the studies that do exist do not provide a clear picture with regard to clinical services. Rasinski, Willis, Baldwin, Yeh, and Lee (1999) found respondents in a private setting responded similarly to sensitive questions whether the format was self-administered or interviewer-administered. Yet other research indicates that anonymity leads to disinhibition and more disclosure, such as when therapy services are provided using non-visually based technology (Richards, 2009). However, the information sought was not evaluative in nature but focused on self-disclosure. In a study focused on the evaluation of clinical services (Soelling & Newell, 1983), client satisfaction scores for treatment were significantly lower when anonymous. Comparisons to these studies are problematic, given that the studies are not evaluating the same constructs.

Even if the condition of anonymity does not influence therapeutic alliance scores, this does not remove concern about the measures themselves or other demand characteristics (e.g., introduction of the measures, likeability of the therapist, vulnerability of the client) that may lead to generally high scores on alliance measures. For example, Hatcher and Gillaspy (2006) noted that clients and therapists may have difficulty being able to differentiate the points at the lower end of alliance measures. Future research should focus on these aspects. One possibility is to utilize a qualitative design and interview clients who rate the alliance at various

levels and to identify conceptual anchors. Do clients conceptualize and articulate the differences in therapeutic alliance similarly to the authors' intention as reflected by the scale of measurement?

The second major finding is that alliance scores were not significantly correlated with a measure of social desirability. In other words, it does not appear that clients' general desire to present favorably was associated with their ratings of the alliance. The concern for social desirability has been noted as a concern by previous researchers (e.g., Tryon et al., 2007) as a potential reason for inflated alliance scores, but the current findings cast doubt on this possibility. There is some, albeit limited, support for this finding in a study by Gaston and Sabourin (1992) that found no correlation between client satisfaction and social desirability when clients completed the measures anonymously.

Another possible explanation is that most clients do actually have a positive perception of the therapeutic alliance; therefore, the scores are accurate (Tryon et al., 2007). If so, concerns regarding demand characteristics and social desirability are diminished–it is easier to give positive feedback. For the current study, most clients rated alliance high. Future research should investigate those individuals who do not have a positive perception of the therapeutic alliance: Are demand characteristics and social desirability then heightened subsequently leading to inflated scores?

The third major finding is that the rate of change in the therapeutic alliance for each of the treatment conditions was similar. All three conditions yielded statistically significant therapeutic alliance gains across the three sessions of treatment, but there was not a statistically significant difference between treatment conditions. SRS score increases for the first three sessions were not influenced by the presence of the therapist or giving feedback to the therapist. This result was also interesting from the perspective that active utilization of alliance feedback did not influence SRS scores. For example, a study by Flückiger et al. (2011) indicated that giving clients brief instructions to be proactive regarding the therapeutic alliance led to increases in subsequent session scores. It would follow that actively discussing SRS scores in session might lead to similar increases but this did not occur. Perhaps simply filling out the measure served as a prompt for clients; it is also important to note that all the conditions completed and discussed the ORS at the beginning of each session, which may have facilitated collaborative discussion in a manner similar to the SRS. Such a process may have increased SRS scores, which contributed to a ceiling effect that limited the ability to detect differences between conditions.

A secondary finding is that the SRS had moderate correlations with the WAI-SR with correlation coefficients ranging from $r = .57$ to $r = .65$, suggesting evidence of concurrent validity for the SRS. This is important to recognize given that the SRS is very brief–only four items. The correlation coefficients are consistent with previous research evaluating the psychometric properties of the SRS (Campbell & Hemsley, 2009; Duncan et al., 2003), and when coupled with the high internal consistency estimates indicate the SRS can be a useful measure for evaluating the therapeutic alliance.

Limitations and Conclusions

There are five main weaknesses of the current study that merit mentioning. First, there was little ethnic/racial diversity in the client sample. Having more variability would have allowed for greater generalizability as well as the possibility to evaluate how ethnicity/race may have moderated the study's findings. Given that racial/ethnic minority clients terminate therapy at a higher rate (Ponterotto, Suzuki, Casas, & Alexander, 2009) and client feedback systems were designed to reduce premature terminations, future feedback research should include more diverse samples.

A second weakness was the absence of a system to monitor intervention, particularly for those who were assigned to complete the measures in the anonymous or delayed feedback conditions. All therapists were trained on how to introduce the measures and had a script to follow for using the measures. However, we are uncertain how rigorously these procedures were followed. Future research could examine how the measures are introduced and whether protocol

deviations matter. Does introducing the measures carefully and encouraging honesty mitigate demand characteristics? Moreover, we are not sure how much therapists focused on the alliance across groups. For example, therapists in the NS Feedback condition may have actually spent more time discussing the alliance because it occurred at the beginning of the next session and was not subject to the same time constraints as discussing the alliance feedback at the end of a session.

Third, we did not have a manipulation check from the clients' perspective. In particular, it is unclear whether participants in the NF condition believed that their therapists would have access to the measure. To emphasize that SRS ratings were anonymous for the NF condition, clients were informed their therapist would not have access to the results, and were given an envelope to seal and instructed to physically place the envelope in a drop-box outside of the therapist's office (two counseling center sites) or in the mailbox of the first author (training site). A fourth limitation was that most clients rated the alliance high, which may have limited our ability to detect differences between the three conditions. As mentioned earlier, this is a comment lament of the alliance research. A fifth limitation of the study is that the social desirability measure (SDS) generated scores of moderate internal consistency ($\alpha = .64$). This concern is tempered by the lack of differences found when considering the demand characteristics of giving feedback directly to the therapist versus in an anonymous manner.

Notwithstanding these limitations, our findings have important implications for therapy research and practice. The current study indicated that concerns for clients being unduly influenced by demand characteristics or social desirability when giving therapists feedback on the therapeutic alliance are perhaps not accurate. We did not find that clients rated the alliance differently when providing this feedback anonymously or in-person to the therapist. Specific to the SRS and the client feedback process, administering the SRS as consistent with the PCOMS protocol yielded scores that were reliable and demonstrated moderate concurrent validity. This is particularly relevant for clinical practice given that monitoring treatment outcome is likely to grow, in part, due to an increased call to demonstrate effectiveness for both clinical training and practice (APA Commission on Accreditation, 2011). For example, the APA Presidential Task Force on Evidence Based Practice (2006) recommended the use of monitoring treatment–seeking feedback regarding the therapeutic alliance is consistent with that recommendation. Research in this area should continue to help ensure that the measures used to monitor treatment and the alliance produce scores that are valid and useful for clinicians and their clients.

References

Ackerman, S. J., Benjamin, L. S., Beutler, L. E., Gelso, C. J., Goldfried, M. R., Hill, C., . . . Rainer, J. (2001). Empirically supported therapy relationships: Conclusions and recommendations of the Division 29 Task Force. Psychotherapy: Theory, Research, Practice, Training, 38, 495–497. doi:10.1037/0033-3204.38.4.495

American Psychological Association Commission on Accreditation. (2011). Commissionon Accreditation Implementing Regulations. Retrieved from http://www.apa.org/ed/accreditation/about/policies/implementing-guidelines.pdf

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. American Psychologist, 61, 271–285. doi:10.1037/0003-066X.61.4.271

Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. Journal of Consulting and Clinical Psychology, 77(4), 693–704. doi:10.1037/a0016062

Anker, M. G., Owen, J., Duncan, B. L., & Sparks, J. A. (2010). The alliance in couple therapy. Journal of Consulting and Clinical Psychology, 78, 635–645. doi:10.1037/a0020051

Ballard, R. (1992). Short forms of the Marlowe-Crowne Social Desirability Scale. Psychological Reports, 71, 1155–1160. doi:10.2466/PR0.71.8.1155-1160

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., . . . McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. Journal of Consulting and Clinical Psychology, 69, 184–196. doi:10.1037//0022-006X.69.2.184

Bordin, E. (1979). The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research, and Practice, 16, 252–260. doi:10.1037/h0085885

Crits-Christoph, P., Gibbons, M., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. Journal of Consulting and Clinical Psychology, 79(3), 267–278. doi:10.1037/a0023668

Crits-Christoph, P., Ring-Kurtz, S., McClure, B., Temes, C., Kulaga, A., Gallop, R., & . . . Rotrosen, J. (2010). A randomized controlled study of a web-based performance improvement system for substance abuse treatment providers. Journal of Substance Abuse Treatment, 38(3), 251–262. doi:10.1016/j.jsat.2010.01.001

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 24, 349–354. doi:10.1037/h0047358

Duncan, B. L. (2010). On becoming a better therapist. Washington, DC: American Psychological Association. doi:10.1037/12080-001

Duncan, B. L., Miller, S. D., & Sparks, J. (2004). The heroic client: A revolutionary way to improve effectiveness through client directed outcome informed therapy (2nd ed.). San Francisco, CA: Jossey-Bass.

Duncan, B., Miller, S., Sparks, J., Claud, D., Reynolds, L., Brown, J., & Johnson, L. (2003). The session rating scale: Preliminary psychometric properties of a "working alliance" inventory. Journal of Brief Therapy, 3(1), 3–11.

Flückiger, C., Del Re, A. C., Wampold, B. E., Znoj, H., & Caspar, F., & Jörg, U. (2011). Valuing clients' perspective and the effects on the therapeutic alliance: A randomized controlled study of an adjunctive instruction. Journal of Counseling Psychology. doi:10.1037/a0023648

Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and empirical considerations. Psychotherapy: Theory, Research, Practice, Training, 27(2), 143–153. doi:10.1037/0033-3204.27.2.143

Gaston, L., & Sabourin, S. (1992). Client satisfaction and social desirability in psychotherapy. Evaluation & Program Planning, 15(3), 227–231. doi:10.1016/0149-7189(92)90085-9

Gillaspy, J. A., & Murphy, J. J. (2011). Incorporating outcome and session rating scales in Solution-focused Brief Therapy. In C. Franklin, T. Trepper, W. Gingerich, & E. McCollum (Eds.), Solution-focused brief therapy. New York, NY: Oxford University Press.

Harmon, S. C., Lambert, M. J., Smart, D. W., Hawkins, E. J., Nielson, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist/client feedback and clinical support tools. Psychotherapy Research, 17, 379–392. doi:10.1080/10503300600702331

Hatcher, R. L., & Barends, A. W. (1996). Patient's view of psychotherapy: Exploratory factor analysis of three alliance measures. Journal of Consulting and Clinical Psychology, 64, 1326–1336. doi:10.1031/0022-006X.64.6.1326

Hatcher, R., & Gillaspy, J. (2006). Development and validation of a revised short version of the Working Alliance Inventory. Psychotherapy Research, 16, 12–25. doi:10.1080/10503300500352500

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K., & Tuttle, K. (2004). The effects of providing patient progress information to therapists and patients. Psychotherapy Research, 14, 308–327. doi:10.1093/ptr/kph027

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. Psychotherapy, 48(1), 9–16. doi:10.1037/a0022186

Horvath, A., & Greenberg, L. (1989). Development and validation of the working alliance inventory. Journal of Counseling Psychology, 36(2), 223–233. doi:10.1037//0022-0167.36.2.223

Kazdin, A. E. (2007). Systematic evaluation to improve the quality of patient care: From hope to hopeful. Pragmatic Case Studies in Psychotherapy, 3, 37–49.

Knoblock-Fedders, L. M., Pinsof, W., & Mann, B. J. (2007) Therapeutic alliance and treatment progress in couple therapy. Journal of Marital & Family Therapy, 33, 245–257. doi:10.1111/j.1752-0606.2007.00019

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., . . . Reisinger, C. (1996). Administration and scoring manual for the OQ 45.2. Stevenson, MD: American Professional Credentialing Services.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., . . . Burlingame, G. M. (2004b). Administration and scoring manual for the Outcome Questionnaire-45. Salt Lake City, UT: OQMeasures.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. Psychotherapy, 48(1), 72–79. doi:10.1037/a0022238

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on client progress during psychotherapy: Are outcomes enhanced? Psychotherapy Research, 11, 49–68. doi:10.1080/713663852

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. K. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. Clinical Psychology and Psychotherapy, 9, 91–103. doi:10.1002/cpp.324

Leary, M. R. (2001) Introduction to behavioral research methods (3rd ed.). Needham Heights, MA: Allyn and Bacon.

Loo, W., & Loewen, P. (2004). Confirmatory factor analysis of scores from full and short versions of the Marlowe-Crowne Social Desirability Scale. Journal of Applied Social Psychology, 34, 2343–2352. doi:10.1111/j.1559-1816.2004.tb01980.x

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-Crowne Social Desirability Scale. Journal of Social Psychology, 140, 628–635. doi:10.1080/00224540009600503

Luborsky, L., Barber, J., Siqueland, L., Johnson, S., Najavits, L., Frank, A., & Daley, D. (1996). The revised Helping Alliance questionnaire (HAQ-II): Psychometric properties. The Journal of Psychotherapy Practice and Research, 5, 260–271.

Miller, S., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. Journal of Brief Therapy, 2(2), 91–100.

Miller, S. D., & Duncan, B. L. (2004). The Outcome and Session Rating Scales: Administration and scoring manuals. Chicago, IL: Author.

Miller, S. D., Duncan, B. L., & Johnson, L. D. (2000). The Session Rating Scale 3.0. Chicago, IL: Authors.

Munder, T., Wilmers, F., Leonhart, R., Linster, H., & Barth, J. (2010). Working Alliance Inventory-Short Revised (WAI-SR): Psychometric properties in outpatients and inpatients. Clinical Psychology & Psychotherapy, 17(3), 231–239. doi:10.1002/cpp.658

Peugh, J. L. (2010). A practical guide to multilevel modeling. Journal of School Psychology, 48, 85–112. doi:10.1016/j.jsp.2009.09.002

Ponterotto, J. G., Suzuki, L. A., Casas, J. M., & Alexander, C. M. (2010). Handbook of multicultural counseling (3rd ed). Thousand Oaks, CA: Sage Publications.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. Applied Cognitive Psychology, 13(5), 465–484. doi:10.1002/(SICI)1099-0720(199910)13:5<465::AID-ACP609>3.0.CO;2-Y

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2005). HLM 6: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International, Inc.

Reese, R., Norsworthy, L., & Rowlands, S. (2009). Does a continuous feedback system improve psychotherapy outcome? Psychotherapy: Theory, Research, Practice, Training, 46, 418–431. doi:10.1037/a0017901

Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. Psychotherapy: Theory, Research, Practice, Training, 47, 616–630. doi:10.1037/a0021182

Richard, D. (2009). Features and benefits of online counselling: Trinity College online mental health community. British Journal of Guidance and Counselling, 37, 231–242. doi:10.1080/03069880902956975

Shimokawa, K., Lambert, M., & Smart, D. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. Journal of Consulting and Clinical Psychology, 78(3), 298–311. doi:10.1037/a0019247

Slade, K., Lambert, M. J., Harmon, S., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. Clinical Psychology & Psychotherapy, 15(5), 287–303. doi:10.1002/cpp.594

Soelling, M. E., & Newell, T. G. (1983). Effects of anonymity and experimenter demand on client satisfaction with mental health services. Evaluation and Program Planning, 6(3–4), 329–333. doi:10.1016/0149-7189(83)90012-5

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. Psychological Assessment, 1, 207–210. doi:10.1037/1040-3590.1.3.207

Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta-analytic examination of client-therapist perspectives of the working alliance. Psychotherapy Research, 17, 629–642. doi:10.1080/10503300701320611

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 failure and problem solving strategies in routine practice. Journal of Counseling Psychology, 58, 59–68. doi:10.1037/0022-0167.50.1.59