Journal of Consulting and Clinical Psychology
2009, Vol. 77, No. 4, 693–704

0022-006X/09/$12.00 DOI: 10.1037/a0016062

Using Client Feedback to Improve Couple Therapy Outcomes: A Randomized Clinical Trial in a Naturalistic Setting

Morten G. Anker
Vestfold Counseling Office of Family Affairs

Barry L. Duncan
Institute for the Study of Therapeutic Change

Jacqueline A. Sparks
University of Rhode Island

Despite the overall efficacy of psychotherapy, dropouts are substantial, many clients do not benefit, therapists vary in effectiveness, and there may be a crisis of confidence among consumers. A research paradigm called *patient-focused research*—a method of enhancing outcome via continuous progress feedback—holds promise to address these problems. Although feedback has been demonstrated to improve individual psychotherapy outcomes, no studies have examined couple therapy. The current study investigated the effects of providing treatment progress and alliance information to both clients and therapists during couple therapy. Outpatients ($N = 410$) at a community family counseling clinic were randomly assigned to 1 of 2 groups: treatment as usual (TAU) or feedback. Couples in the feedback condition demonstrated significantly greater improvement than those in the TAU condition at posttreatment, achieved nearly 4 times the rate of clinically significant change, and maintained a significant advantage on the primary measure at 6-month follow-up while attaining a significantly lower rate of separation or divorce. Mounting evidence of feedback effects with different measures and populations suggests that the time for routine tracking of client progress has arrived.

Keywords: patient-focused research, couple therapy, practice-based evidence, routine client-based outcome and alliance assessment, feedback

It is often reported that the average treated person is better off than approximately 80% of the untreated sample (Lambert & Ogles, 2004; Wampold, 2001), which translates to an effect size (ES) of about 0.8. These substantial benefits apparently extend from the laboratory to the real world of everyday practice. For example, Minami et al. (2008) found comparable results to those reported in randomized clinical trials (RCT) for the treatment of depression in a managed care population. In short, the good news is that psychotherapy works.

This is, however, a "good-news, bad-news" scenario. First, dropouts are a significant problem in the delivery of mental health services, averaging at least 47% (Wierzbicki & Pekarik, 1993). Second, despite the fact that the general efficacy of psychotherapy is consistently good, not everyone benefits. Hansen, Lambert, and Foreman (2002), using a national database of over 6,000 clients averaging five sessions of psychotherapy, reported a different and sobering picture of routine clinical care in which only 20% of clients improved, compared with the 57%–67% rates typical of RCTs. Whichever rate is accepted as more representative of actual practice, the fact remains that not everyone benefits.

Perhaps explaining part of the wide range of results, variability among therapists continues to be the rule rather than the exception (Beutler et al., 2004). In a study of managed care treatment, Wampold and Brown (2005) reported that 5% of outcome was attributable to therapist variability. Additional studies of routine care have indicated small to moderate therapist effects (Okiishi et al., 2006). Baldwin, Wampold, and Imel (2007) found that the therapist's variability in the alliance predicted outcome, suggesting that the alliance may represent an arena for influencing the variance due to the therapist.

Finally, perhaps as a result of the previous two points, consumer confidence in psychotherapy is troubling. A survey by the American Psychological Association (APA; Penn, Schoen, & Berland Associates, 2004) asked 1,000 potential consumers, "Is this an important reason why you might choose not to seek help from a mental health professional?" The highest percentage of responses were lack of insurance (87%) and concerns about cost (81%). The third most endorsed reason was a lack of confidence in the outcome of treatment (77%). So despite the overall efficacy and effectiveness of psychotherapy, dropouts are a substantial problem, many clients do not benefit, therapists vary significantly in effectiveness, and there seems to be a crisis of confidence among consumers.

A relatively new research paradigm called *patient-focused research* (Howard, Moras, Brill, Martinovich, & Lutz, 1996) or

Morten G. Anker, Vestfold Counseling Office of Family Affairs, Regional Office for Children, Youth and Family Affairs, Tønsberg, Norway; Barry L. Duncan, Institute for the Study of Therapeutic Change, Tamarac, FL; Jacqueline A. Sparks, Department of Human Development and Family Studies, University of Rhode Island.

We thank Bruce Wampold and Zac Imel for invaluable statistical consultation and analysis.

Correspondence concerning this article should be addressed to Barry Duncan, Institute for the Study of Therapeutic Change, 8611 Banyan Court, Tamarac, FL 33321. E-mail: barrylduncan@comcast.net

practice-based evidence (Barkham et al., 2001) holds great promise to address these problems. Howard et al. (1996) advocated for the systematic evaluation of client response to treatment during the course of therapy and recommended that such information be used to "determine the appropriateness of the current treatment . . . [and] the need for further treatment . . . [and] prompt a clinical consultation for patients who [were] not progressing at expected rates" (Howard et al., 1996, p. 1063). Although several quality assurance systems seek to enhance outcome via continuous monitoring and feedback to clinicians (see Lambert, in press), the pioneering work of Michael Lambert and colleagues stands out—not only for the development of measurement systems, statistical methodologies, and predictive algorithms, but also for their groundbreaking investigations of the effects of providing therapists feedback about client progress in treatment.

In a meta-analysis of three trials (Lambert et al., 2001, 2002; Whipple et al., 2003), Lambert et al. (2003) reported an ES of 0.39 for feedback when comparing the gains of clients identified as deteriorating who were in the feedback group (therapists were provided feedback) with the treatment as usual (TAU) or nonfeedback group. Studies by Whipple et al. (2003) and Harmon et al. (2007) found that adding measures of the alliance, motivation to change, and perceived social support for clients identified as not on track demonstrated incremental effectiveness over the continuous feedback model alone. Two studies looked at whether providing feedback to both therapist and client influences effectiveness. Hawkins, Lambert, Vermeersch, Slade, and Tuttle (2004) found that giving feedback on progress to both clients and therapists was associated with significant gains in outcome. However, Harmon et al. (2007) failed to replicate these results.

All five studies shared design features that strengthen the case for tracking client progress: (a) random assignment, (b) the use of the same therapist across treatment conditions, (c) a variety of treatment approaches, and (d) a high percentage of licensed clinicians. All five trials realized significant gains for feedback groups over TAU for at-risk clients. Twenty-two percent of TAU at-risk cases reached reliable improvement and clinically significant change, compared with 33% for feedback to therapist groups, 39% for feedback to therapists and clients, and 45% when feedback was supplemented with clinical support tools (Lambert, in press). Three of the five studies suggested that feedback enhances outcome for clients who are at risk but yield little impact for other clients (Lambert, in press). Two studies (Harmon et al., 2007; Hawkins et al., 2004) found that using continuous assessment was helpful to all clients, although those who were predicted to not succeed in treatment benefited more. In total, this research makes a strong case for routine measurement of outcome in everyday clinical practice (Lambert, in press).

Another method of using continuous client feedback to improve outcomes is the Partners for Change Outcome Management System (PCOMS; Duncan, Miller, & Sparks, 2004; Miller, Duncan, Sorrell, & Brown, 2005). Much of this system's appeal rests on the brevity of the measures and therefore its feasibility for everyday use in the demanding schedules of front-line clinicians. The Outcome Rating Scale (ORS) and the Session Rating Scale (SRS; Miller & Duncan, 2004) are both four-item measures developed to track outcome and the therapeutic alliance, respectively. PCOMS was based on Lambert et al.'s (1996) continuous assessment model using the Outcome Questionnaire 45, but there are differences beyond the measures. First, PCOMS is integrated into the ongoing psychotherapy process and routinely includes a transparent discussion of the feedback with the client (Duncan et al., 2004). Session by session interaction is focused by client feedback about the benefits of psychotherapy or lack thereof. Second, PCOMS assesses the therapeutic alliance in every session and includes a discussion of any potential problems. Lambert's system includes alliance assessment only when there is a lack of progress.

Miller, Duncan, Brown, Sorrell, and Chalk (2006) used PCOMS to explore the impact of feedback in a large, culturally diverse sample that was using telephonic employee assistance services. Whereas the study's quasi-experimental design—baseline scores obtained in Phase 1 provided benchmarks for later phase comparisons—limits the conclusions that can be drawn, results were consistent with those found in RCTs that examined feedback with other measures. The use of outcome feedback doubled the ES from 0.37 to 0.79 and significantly increased retention.

A recent investigation used PCOMS to study the effects of feedback versus TAU (Reese, Norsworthy, & Rowlands, 2008). Findings from two samples of clients who attended individual therapy at a university counseling center or a graduate training clinic demonstrated statistically significant treatment gains for feedback when compared with TAU. Reese at al. (2008) also reported that clients using PCOMS were more likely to experience reliable change than their TAU counterparts and that the effects of continuous feedback extended to all clients in the feedback condition, not just to those at risk for a negative outcome.

Although the above studies support the use of practice-based evidence in individual psychotherapy, no studies have explored feedback in couple therapy. Meta-analyses have demonstrated that couple therapy has a similar proven record of efficacy over no treatment, ranging from an ES of 0.61 (Shadish et al., 1993) to 0.84 (Shadish & Baldwin, 2003). Shadish and Baldwin (2005) meta-analytically examined randomized trials of the most investigated approach, behavioral marital therapy (BMT), and found it significantly more effective than no treatment ($d = 0.59$). In an RCT of 134 couples, Christensen et al. (2004) reported an ES of 0.86 for traditional behavioral couple therapy (TBCT) and integrative behavioral couple therapy (IBCT). Finally, Gollan and Jacobson (2002) identified four couple treatments, in addition to TBCT and IBCT, with proven efficacy over no treatment: emotionally focused couple therapy (EFCT; Greenberg & Johnson, 1988); cognitive-behavioral marital therapy (CBMT; Baucom & Epstein, 1990); strategic therapy (Goldman & Greenberg, 1992); and insight-oriented marital therapy (IOMT; Snyder & Wills, 1989).

Estimates vary regarding the power of couple therapy to produce clinically significant change (Jacobson & Truax, 1991). Originally, Jacobson et al. (1984) reported a 35.5% success rate, although Shadish et al. (1993) calculated that 41% of the treatment conditions in their review brought couples from a distressed to a nondistressed status. More recently, Shadish and Baldwin (2003) suggested that between 40% and 50% of couples were treated successfully. Confirming this conclusion, the Christensen et al. (2004) trial found that 48% of couples reached recovered status on the Dyadic Adjustment Scale (Spanier, 1976). Similar to individual psychotherapy, however, effectiveness may be substantially less in actual clinical settings, especially considering dropouts. For example, Hahlweg and Klann (1997) reported an ES of 0.28 for couple clinicians in Germany and a 49% attrition rate.

Differential efficacy among couple approaches has yet to be established. Dunn and Schwebel's (1995) meta-analysis of BMT, CBMT, IOMT, and EFCT reported that weighted mean effect sizes were not significantly different at either posttreatment or follow-up on marital behavior, including target complaint. IOMT was significantly better on relationship ratings at posttreatment but not at follow-up. In a review of 20 meta-analyses of couple and family interventions, Shadish and Baldwin (2003) found few significant differences between various models. In their large comparison of TBCT and IBCT, Christensen et al. (2004) reported, "For the most part, TBCT and IBCT performed similarly across measures, despite being demonstrably different treatments" (p. 188). The lack of reliable superiority of any particular couple treatment suggests that improving outcomes may require a different methodology than transporting specific models to clinical settings.

The current study uses PCOMS, a method not tied to a single orientation, to examine the impact of feedback in couples work. Using a randomized design within routine care and following the design features found in Lambert's studies, TAU (no feedback) is compared with a feedback condition in which both therapists and couples have access to client-generated alliance and outcome information at each session. Specifically, the study seeks to answer how outcomes for couples and therapists receiving systematic feedback on progress and the alliance differ from those not receiving feedback at posttreatment and 6-month follow-up. If there is a differential effect, is it limited to couples identified as not on track and likely to deteriorate, or is there a more general beneficial effect for the feedback condition? It is hypothesized that couples who receive systematic feedback will have significantly better outcomes than nonfeedback couples. The study secondarily examines therapist variance in work with couples. It is unclear how therapist and feedback variables interact and what are their relative contributions and interdependence. Finally, the current study seeks to address concerns that couple trials do not generalize well into everyday practice (Shadish & Baldwin, 2003). Conducted in a naturalistic site with highly feasible instruments, the study asks how outcomes for couples therapists using diverse treatment approaches in everyday practice are impacted by the routine incorporation of client feedback.

Method

Participants

A total of 906 individuals (453 couples) who sought outpatient couple therapy services at a family counseling agency providing free government-subsidized services in southern Norway from October 2005 to December 2007 were randomized to one of two groups following phone intake: TAU and feedback. Seventy-seven couples failed to meet the inclusion criteria. Couples were excluded at phone intake when one member refused to attend, one or both members of the couple expressed the desire to end the relationship, or one or both refused informed consent. One hundred thirty-four couples (29.6%) did not attend the first session. This no-show–cancelation rate is about average for this clinical context. The reasons for cancellation or no-show are unknown because confidentiality requirements prohibit the collection of data on those who do not attend the first session. In the final sample, couples were required to have attended at least two sessions of treatment and completed the outcome measure for a minimum of the first and last session, which eliminated another 30 couples. Two hundred five couples (410 individuals) completed pre- and posttreatment measures—102 couples in the TAU condition (204 individuals) and 103 couples (206 individuals) in the feedback group.

The age of the individuals in the final sample ranged from 20 to 71 years with a mean of 37.83 (SD = 8.48). Three hundred and sixteen (77.1%) participants were employed full-time and 34 (8.3%) were employed part time, whereas 60 (14.6%) were unemployed or did not work outside the home. Regarding education levels, 63 (15.4%) had completed lower secondary school, 186 (45.4%) had completed upper secondary school, and 161 (39.3%) had completed university or college. The mean number of years the couples had been together was 11.2 (SD = 8.2). Before the first session, study participants were also asked to identify their goals for therapy on a standard intake form. Two hundred ninety-eight (72.6%) participants marked the goal of achieving a better relationship, whereas 98 (23.9) sought clarification regarding whether the relationship should continue. Six individuals (1.5%) indicated a goal of terminating the relationship in the best possible way, and another 6 (1.5%) marked *other* without elaboration.

Couples were white, Euro-Scandinavian, and heterosexual. Couples self-referred with a broad range of typical relationship problems, including communication difficulties (84 couples), loss of feeling for partner (37), jealousy or infidelity (32), conflict (30), and coping with partner's physical or psychological problem (22). Diagnosis was not required, nor is it a routine convention in this setting. The mean intake score of the 410 participants on the ORS (Miller, Duncan, Brown, Sparks, & Claud, 2003; see below) was 18.33 (SD = 7.45), indicative of a clinical population and similar to distress levels of other clinical sites (Miller & Duncan, 2004). Similarly, the mean marital satisfaction score on the Locke-Wallace Marital Adjustment Test (LW; Locke & Wallace, 1959; see below) was 72.11 (SD = 24.73), indicative of a dissatisfied relationship and well under the traditional cutoff score of 100.

Follow-Up Participants

A total of 245 (59.8%) out of 410 individuals, representing 149 couples, responded to 6-month follow-up. In the follow-up sample, like the pre–post sample, the couples were required to have full data sets from both individuals for inclusion: They had attended at least two sessions of treatment and completed both outcome measures for the first and follow-up evaluations. This eliminated 97 individuals, 53 (from 39 couples) of whom were divorced or separated. Seventy-four couples (148 individuals) completed pretreatment and follow-up measures: 32 in the TAU condition (64 individuals) and 42 (84 individuals) in the feedback group. The mean ORS score at pretreatment was 19.45 (SD = 7.72). The mean marital satisfaction score on the LW was 78.28 (SD = 24.40). Although higher than the total sample, both measures indicate a pretreatment clinical population (see below). The total follow-up sample of 149 couples was used to calculate the separation or divorce rate at follow up.

Therapists

The couples were seen by 10 therapists (7 female and 3 male). Four were licensed psychologists, 5 were licensed social workers,

and 1 was a licensed psychiatric nurse. All therapists professed an eclectic orientation, using a variety of approaches—solution-focused, narrative, cognitive–behavioral, humanistic, and systemic—similar to those typically practiced in Norway family counseling agencies. Because this investigation was intended to reflect couple therapy practice in typical clinical settings, consistent with effectiveness rather than efficacy methodology, checks were not performed to ensure treatment integrity.

The average age of the therapists was 42 years (SD = 13.0 years), and the mean years of experience with couple therapy was 5.1 years (SD = 6.3 years). The number of couples treated by each therapist ranged from 4 to 47 (4, 11, 12, 12, 15, 20, 24, 26, 34, 47). The therapist with 4 couples left the agency; 3 therapists were part time, and 6 were full time. The therapist with the most couples had greater availability during the study.

A simple attitude survey developed for this study was administered to determine therapists' views about attaining client feedback via assessment instruments. None of the therapists were experienced in assessing client progress, and all believed that their usual methods of acquiring feedback (asking clients and evaluating by impression) would be as effective. Therapists held attitudes about continuous assessment that ranged from neutral (four therapists) to positive (six therapists).

Measures of Progress and Outcome

The ORS. Psychological functioning and distress was assessed pre- and posttreatment and at follow-up using the ORS (Miller & Duncan, 2004), a self-report instrument designed to measure client progress repeatedly (at the beginning of each session) throughout the course of therapy. The ORS is a 4-item visual analog scale providing a total score (40) based on 4 subscale domain scores (each with a possible score ranging from 0 to 10) that reflect key areas of client functioning: *individually* (personal well-being), *interpersonally* (family, close relationships), *socially* (work, school, friendships), and *overall* (general sense of well-being). Clients put a mark on the line of each item nearest the pole that best describes their experience, and therapists score each 10-cm line using a centimeter ruler. The scores are totaled, ranging from 0 to 40. Lower scores reflect more severe distress. The ORS total score, a global assessment of client functioning, was used in the current study to provide the measure of change from session to session on which feedback to therapists and clients was based, as well as the criterion measure for ultimate outcome.

Miller et al. (2003) reported that the internal consistency of the ORS was .93 and test–retest reliability was .66. In the current sample, the internal consistency of the ORS was .83. Concurrent validity of the ORS has been demonstrated as adequate through correlates with the Outcome Questionnaire 45 (Lambert et al., 1996; r = .59; Miller et al., 2003), the Symptom Checklist-90-Revised (Derogatis, 1992; r = .57; Reese et al., 2008), and the Clinical Outcomes in Routine Evaluation (Barkham et al., 2001; r = .67; Miller & Duncan, 2004). The ORS represents an attempt to balance the reliability and validity of longer assessment tools with the feasibility required for routine clinical practice (Miller et al., 2003).

The Locke-Wallace (LW) Marital Adjustment Test. The LW (Locke & Wallace, 1959) is a 15-item self-report measure covering domains of marital functioning. The LW has enjoyed broad use and is considered a reliable and valid measure of marital satisfaction, still relevant to clinical practice and research (Freeston & Plechaty, 1997). The LW is highly correlated with the oft-used Dyadic Adjustment Scale (.93; Spanier, 1976). The LW cut-off score of 100, which differentiates satisfied from dissatisfied couples, is widely accepted (Christensen et al., 2004; Freeston & Plechaty, 1997). In the current study, the alpha for the LW was .71. The LW was administered at pretreatment and at the 6-month follow-up.

Clinical significance and reliable change. Using formulas developed by Jacobson and Truax (1991), clinical and normative data for the ORS were analyzed by Miller and Duncan (2004) to provide cutoff scores for the reliable change index and clinically significant change. Using a sample of 34,790 participants, clients who changed in a positive or negative (deterioration) direction by at least 5 points were regarded as having made reliable change.

This degree of change exceeds measurement error, based on the reliability of the ORS, and is one of the two criteria posited by Jacobson and Truax (1991) as indicative of clinically meaningful change. The second criterion requires movement from a score typical of a clinical population to one typical of a functional population. On the ORS, the cutoff at which a person's score is more likely to come from a dysfunctional than a nondysfunctional population is 25 (Miller et al., 2003).

Design and Procedure

This was a pragmatic study conducted in a typical community-based outpatient setting rather than a research setting. The simple randomization procedure (intake forms were shuffled, and then a coin flip determined assignment to the feedback vs. TAU groups) occurred after initial intake notes were taken, but before therapists were assigned cases. After randomization to the treatment groups, which were never changed, clients were invited to participate in a research study about improving the benefits of therapy. All participating clients gave their informed consent, and institutional review and approval was secured. Clients were not informed about the different conditions of feedback and TAU. Participant intake forms were then assigned weekly to available therapist intake slots, two at a time, one from feedback and one from TAU. Therapists could exchange one case for another of the same group only (feedback or TAU) if they felt uncomfortable with the couple's clinical presentation as depicted on the intake paperwork or had any previous nonclinical contact with the couple. Such an exchange happened 10 times over the course of the study, primarily because of previous nontherapy contact with the couple. Therapists in the study were informed that the purpose of the study was to test the effects of feedback in couple therapy. All therapists worked simultaneously with couples from the feedback and TAU groups, with 50% of their study caseload from each. Figure 1 depicts the flow of the clients into the randomized groups.

All therapists attended 2 days of training (8 hr total) before the study and three 3-hr follow-up training sessions during the investigation. Training included the rationale for continuous assessment—that is, that client's subjective experience of early change and the alliance are reliable predictors of ultimate treatment outcome (Haas, Hill, Lambert, & Morrell, 2002; Martin, Garske, & Davis, 2000). Therapists were instructed to follow the general protocol outlined in the scoring and administration manual for the

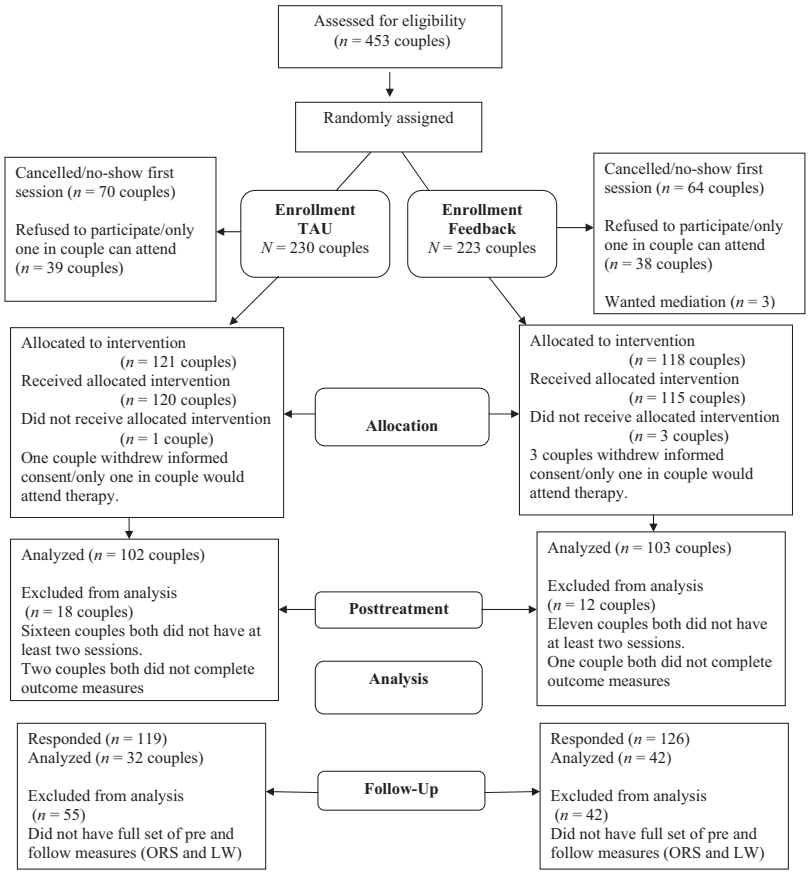

Figure 1. Participant flow into treatment conditions and data analysis.

ORS and SRS (Miller & Duncan, 2004), as well as the transparent, collaborative process of monitoring outcome with clients described in these authors' other publications (e.g., Duncan et al., 2004). The client completes the ORS at the beginning of every session; then, the therapist scores the items with a centimeter ruler and totals the subscale scores. The total ORS score is then charted on a provided graph that indicates a client's progress, or lack thereof, across the course of treatment. The scores can be used to frame content for a session, although the therapist has discretion over how to best integrate the scores. Miller and Duncan (2004) recommended the following general guidelines:

> Deteriorating: Clients dropping 5 points are at risk for drop out or poor outcome. Discuss possible reasons, review the alliance, and consider changing the treatment approach or therapist if deterioration is not quickly abated.
>
> No Change: Clients not showing a reliable change after three sessions are at risk for drop out or poor outcome. Review the alliance, and consider changing the treatment approach. If there is no reliable change after six sessions, seek consultation, supervision, or referral options.
>
> Reliable Change: For clients showing a gain of 5 points, reinforce and consolidate changes until progress begins to plateau, whereupon reducing the frequency of sessions should be discussed.
>
> Clinically Significant Change: For clients showing a change of 5 points and have crossed the cut off, consolidate changes, anticipate potential setbacks, and consider a reduction of the frequency of sessions or termination.

Therapists were also instructed in the use of the SRS, a 4-item visual analogue scale (Duncan et al., 2003) used to detect potential breaches in the alliance. Toward the end of every session, the client completes the SRS and the therapist scores it; this allows the therapist to openly discuss any concerns and how the services may better fit the client's expectations. The total score is then charted on the same graph as the ORS. In the current study, the SRS was used as part of the PCOMS feedback process but not included in the analysis.

In addition, therapists were trained on how to integrate ORS feedback using a table of expected treatment responses (ETR). On the basis of the intake score, using algorithms derived from the ORS database of over 300,000 administrations, a web-based program calculates trajectories of change at 25th, 50th, and 75th percentile levels. A given client achieved the ETR when his or her score met or exceeded the 50th percentile trajectory for all clients in the database who had entered with the same intake score. Clients were determined to be at risk when their ORS scores fell below the 50th percentile at the third session (or the second session, if the couple only attended two sessions). The ETRs were based on individual responses to treatment and were not specific to couple therapy.

Therapists were advised to initiate a discussion with couples in the feedback condition if one or both individuals of the couple were not on track or were at risk (i.e., individuals or couples who fell below the 50th percentile trajectory). Given that the training emphasized the reliable predictors of treatment success (i.e., early change and the alliance), it is worth noting that this training may have enhanced outcomes in both conditions; the training could have diminished the effects of feedback by increasing therapist attention in the TAU condition to early change and the alliance.

Feedback and TAU. Therapists in the feedback condition followed the procedure discussed above. Two simple ways to incorporate and discuss feedback with clients were used. Therapists were given graphs to plot ORS and SRS scores and were encouraged to show the graph to clients and discuss its ongoing implications Therapists were also given a table of 50th percentile ETRs for all possible intake scores (0–40), enabling them to compare their clients' ongoing scores with the 50th percentile expected change trajectory derived from the ORS database. To ensure that therapists used the table for comparison at least once, therapists were asked to put the expected change score from the second session on the couple's graph. The primary investigator reviewed charts weekly. Therapists and clients, therefore, had ongoing access to ORS and SRS scores and ETRs for clients in the feedback condition. Although the procedures of this study strongly encouraged therapists to show the graphs and comparisons to clients and to openly discuss the feedback with them, the frequency or content of these interactions was not monitored. In addition, no attempt was made to influence a particular response on the basis of feedback. Therapists were free to devise the same or different treatment strategies as they deemed appropriate.

The clients in the TAU condition received the ORS from the secretary and filled it out before the first session and all subsequent sessions. TAU client scores were placed in sealed envelopes and were not accessible to the therapists.

Follow-up. Six months after the last session, each participant was mailed a packet containing a prepaid addressed envelope, the LW, ORS, and other questions regarding their experiences in treatment, including whether the couple remained together or were separated or divorced. If no response was received within 3 weeks, another packet was sent.

Statistical Analyses

Multilevel models were used to test hypotheses, taking into account the nested structure of couple therapy data. Multilevel models are advantageous because researchers can model dependencies that are likely to be present in data that have a nested structure, allowing an unbiased estimate of the feedback effect (Atkins, 2005). If some therapists are generally more effective than others, then the outcomes of couples seen by the same therapists will be correlated; as well, the outcomes of individuals within couples are correlated (if one improves, then it may be that the other member of the couple improves), thus violating the independence assumption of traditional statistical tests. The data were structured in three levels as follows: Individuals (Level 1) were nested within couples (Level 2), and couples were nested within therapist (Level 3). Data was analyzed with hierarchical linear modeling (HLM6; Raudenbush, Bryk, Cheong, & Congdon, 2005).

Three multilevel models were constructed to examine therapist and couple effects as well as the effects of feedback on posttreatment functioning. In Model 1, we tested for the presence of therapist and couple effects by controlling for pretreatment functioning at Level 1 and evaluating the residual intraclass correlations. Specifically, it is possible to estimate how much of the variability in outcomes is attributable to the therapist and the couple. To determine the variance due to the therapist and couple, we calculated the residual intraclass (controlling for pretreatment ratings) correlation, ρ_i (Kenny & Judd, 1996). In Model 2, we estimated the effect of feedback with a dichotomous predictor at Level 2, and in Model 3 we added a random slope parameter to examine if the effect of feedback varied across therapists. The final equation is as follows:

$$Y_{ijz} = \beta_{00} + \beta_{10}(\mathrm{ORS}_{ijz}) + \beta_{01}(\mathrm{FEEDBACK}_{jz}) + [R_0 + U_{00} + U_{01} + E],$$

where Y_{ijz} is the posttest ORS score for the client *i,* in couple *j*, treated by therapist *z*; ORS (grand mean centered) is the pretest severity for client *i,* in couple *j,* seen by therapist *z*; β_{00} is the intercept; β_{10} is the regression coefficient for pretest ORS at level 1; R_0 is the between-couple variance ($\sigma^2_{\mathrm{couple}}$); U_0 is the between-therapist variance (σ^2_{ther}); U_{01} provides the variance in the size of β_{01} effect across therapists; and E is the variance at the client level (σ^2_e). Coefficients inside the brackets are the random effects and coefficients outside the brackets are the fixed effects. We replicated these three multilevel models with follow-up data, controlling for pretreatment functioning at Level 1 and entering feedback as predictor of follow-up functioning at Level 2. Additional analyses were also performed to determine the percentage of couples meeting reliable and clinically significant change criteria as well as the percentage of couples at risk who ultimately responded to treatment.

Results

We first examined the sample on several demographic variables to determine if randomization was successful. A series of independent samples *t* tests revealed no evidence of differences between the couples in the feedback and couples in the TAU condition. Specifically, no between-group differences were found on the mean ORS score at pretreatment, $t(410) = 0.69$, $p > .05$; on the age of the clients, $t(410) = 0.07$, $p > .05$; or on years as a couple, $t(410) = 1.02$, $p > .05$. Chi-square analyses also did not reveal any differences in employment status, education, problems, and goals for treatment. No significant differences between the groups were found in the follow-up sample, either.

Pretreatment, posttreatment, and follow-up means and standard deviations for each condition on the ORS and LW (LW only collected at pretreatment and follow-up) are reported in Table 1, as well as mean number of sessions (length of stay). Table 2 provides the test of the effect of feedback and presents the fixed and random effects from the multilevel models for the ORS.

Model 1: Base Model

Model 1 provides the fixed effects for pretreatment scores. The Level 1 coefficient (π_{10}) indicates that the higher an individual's

Table 1
Pretreatmemt, Posttreatment, and Follow-Up Means and Effect Sizes on the Outcome Rating Scale and Locke-Wallace Marital Adjustment Test

	Feedback		Treatment as usual (TAU)	
Time of measurement	ORS *M (SD)*	LW *M (SD)*	ORS *M (SD)*	LW *M (SD)*
Pretreatment score	18.08 (7.85)	78.76 (23.97)	18.58 (7.03)	77.97 (25.51)
Posttreatment score	26.35 (10.02)		21.69 (8.69)	
d (pretreatment–posttreatment)	1.05		0.44	
Posttreatment LOS	4.75 (2.71)		4.45 (2.73)	
Follow-up score	28.28 (9.11)	91.16 (28.48)	24.60 (7.4)	83.06 (27.42)
d (pretreatment–follow-up)	1.14	0.52	0.64	0.21
Follow-up LOS	5.36 (2.97)		4.81 (3.48)	

Note. $N = 410$ (206, feedback; 204, TAU) for pretreatment and posttreatment scores, and $N = 148$ (84, feedback; 64, TAU) at follow-up. Pretreatment–posttreatment and pretreatment–follow-up effect sizes were calculated by dividing the mean difference by the pretreatment standard deviation. ORS = Outcome Rating Scale (Miller & Duncan, 2004); LW = Locke-Wallace Marital Adjustment Test (Locke & Wallace, 1959); LOS = length of stay.

pretreatment score, the higher his or her expected score at posttreatment. Specifically, the predicted posttreatment ORS score of an individual with an average pretreatment ORS was 23.98, indicating that ORS scores improved significantly from pre- to posttreatment (see Table 2). The standardized mean difference of pre- and posttreatment means was $d = 0.76$, which is considered a large ES (Cohen, Cohen, West, & Aiken, 2003) and is comparable to the effects observed in other psychological treatments (see Wampold, 2001).

Examination of the random effects for Model 1 in Table 2 indicates a significant couple effect, but the therapist-level variance component did not reach significance ($p = .074$). The significant couple-level variance component for the ORS indicated there was significant variability among couples in posttreatment scores adjusted for pretreatment. The intraclass correlation (ρ_{couple}) was .50, indicating posttreatment functioning was highly similar within couples. That is, if one partner improved as a result of treatment, then the other partner also improved. The intraclass correlation for therapists (ρ_{ther}) was .02, which is somewhat smaller than typically observed in clinical trials. If the partner level variability is ignored (partners within couples), then the intraclass correlation coefficient is .04, which is in the range of other naturalistic therapist effects (see Baldwin et al., 2007; Wampold & Brown, 2005). That is, about 4% of the variability in outcomes (adjusted for pretest scores), ignoring differences between ratings of couples, was due to the therapist.

Table 2
Multilevel Models Predicting Posttest Outcome Rating Scale

Effects	Model 1 coefficient	Model 2 coefficient	Model 3 coefficient
Fixed effects			
Intercept (γ_{00})	23.99***	21.54***	21.56***
Client ORS (π_{10})	0.43***	0.45***	0.45***
Feedback (β_{01})		4.89***	5.15*
Random effects			
Couple variance (σ^2_{couple})	39.23***	32.70***	28.10***
Therapist variance (σ^2_{ther})	1.52	1.80*	0.43
Client variance (σ^2_e)	37.72***	37.89***	37.93***
Slope of feedback (U_{02})			18.79**
ρ_{couple}	.50	.45	
ρ_{ther}	.02	.03	

Note. ORS = Outcome Rating Scale (Miller & Duncan, 2004).
* $p < .05$. ** $p < .01$. *** $p < .001$.

Model 2: Effects of Feedback

In Model 2, we tested the effect of feedback adding an additional parameter at the couple level (Level 2). The coefficient for feedback (β_{01}) was a significant and positive predictor of ORS scores at posttreatment after controlling for pretreatment functioning. Specifically, the predicted score of an individual in a couple with a therapist who received feedback was 4.89 points higher on the ORS than one who did not (see Table 2). The ES for individuals in couples who received feedback versus those who did not was $d = 0.50$ (calculated by the dividing the mean difference by the pooled standard deviation and correcting for bias). This effect size is in the moderate range for the social sciences (Cohen et al., 2003). It is at least twice as large as the upper bound of the difference between psychological therapies intended to be therapeutic (Benish, Imel, & Wampold, 2008; Wampold, 2001).

Model 3: Differential Effects of Feedback

Finally, in Model 3 we added a random slope parameter, allowing the effect of feedback to vary across therapists. This analysis was a test of a random slope and intercept model. There was significant variability in slopes between therapists in the effect of feedback, indicating that the effect of feedback on posttreatment functioning of clients varied significantly across therapists. An inspection of the empirical Bayes residuals for the feedback effect suggested the variability in feedback was not due to a single outlier therapist. Specifically, the residuals ranged from −6.72 to 7.33, $M = 0.00$, $SD = 3.80$. Due to the small number of therapists ($n = 10$), we did not test hypotheses about which therapists benefited more from feedback than others. However, the correlation between variability in therapist intercepts (variability in the effectiveness of a therapist with no feedback) and variability in the effect of feedback was unusually high, $r = -.99$. Although the small

number of therapists significantly limits any conclusions that can be drawn, it does suggest that the less effective therapists (those who had the worst outcomes without feedback) benefited more from feedback that the most effective therapists. Though caution in interpretation is appropriate, it may be that the effects of feedback are more pronounced for those therapists with poorer outcomes, and therefore the benefits of feedback are exerted by improving the outcomes of less effective therapists. These preliminary findings based on only 10 therapists warrant further investigation and replication. Note that Model 3 did not allow the estimation of an overall interclass correlation (i.e., the intraclass correlation can only be computed for a specific value of the mean pretreatment group severity).

Clinical Significance

To further determine the clinical meaningfulness of treatment gains, final outcomes were categorized according to the number of couples who responded to treatment (met either reliable or clinical significant change criteria) and those who did not respond to treatment (deteriorated or no change). For couples to be considered reliably or clinically significantly changed, both individuals within the couple were required to meet the described criteria. The proportion of clients responding to treatment in the TAU group was 41.7% (both in couple, 22.6%) and in the feedback group was 64.6% (both in couple, 50.5%). A chi-square analysis of the proportion of responding couples revealed a significant difference, $\chi^2(1, N = 205) = 17.24$, $p < .001$. Note that chi-square analyses were not performed on individual data, as assumptions of independence were violated.

Regarding the at-risk or not-on-track couples, chi-square analyses found that significantly fewer at-risk cases emerged in the feedback condition (74.5% in TAU vs. 54.4% in feedback group), $\chi^2(1, N = 205) = 9.07$, $p < .001$. The proportion of at-risk couples who responded (both in couple) in the TAU condition was 9.2% (7 couples) and 28.6% (16 couples) for the feedback condition, a significant difference, $\chi^2(1, N = 132) = 8.4$, $p < .01$. The frequencies and proportions of couples for all the described categories for pretreatment, posttreatment, and follow-up are presented in Table 3.

Follow-Up

The predicted follow-up ORS score of an individual with an average pretreatment ORS was 26.69, indicating that ORS scores increased slightly after termination but were still improved from pretreatment (see Table 4). The standardized mean difference of pretreatment and follow-up ORS was $d = 0.94$. The coefficient for feedback (β_{01}) was significant but was somewhat smaller than the pre- to posttreatment effect. Specifically, the predicted score of an individual with an average ORS score in a couple with a therapist who received feedback was 3.97 points higher on the ORS than one who did not. The ES for individuals from couples who received feedback versus those who did not was $d = 0.44$. More specifically, at 6-month follow-up, the proportion of clients responding to treatment as measured by the ORS in the TAU group was 39.1% (both in couple, 18.8%) and in the feedback group was 66.7% (both in couple, 47.6%). A chi-square analysis of the proportion of responding couples revealed a significant difference, $\chi^2(1, 74) = 6.42$, $p = .01$.

The effects of feedback at follow-up were also assessed by examining the marital status of couples as well as their marital satisfaction, as measured by the LW. At follow-up, a significantly greater proportion of couples were intact (i.e., not divorced or separated) in the feedback condition (81.59%) than in the TAU condition (65.75%), $\chi^2(1, 149) = 4.83$, $p = .014$. The LW was collected only for individuals in intact couples ($n = 148$; one cannot rate satisfaction with a relationship that does not exist), who presumably have more satisfactory relationships than those couples who chose to separate; thus, it would be difficult to find differences between the feedback and TAU conditions. Nevertheless, there was a trend toward greater marital satisfaction in the intact feedback couples vis-à-vis the TAU intact couples. The predicted follow-up LW score of an individual with an average

Table 3
Couples (Both in Couple) and Outcome Classifications on the Outcome Rating Scale at Posttreatment and Follow-Up

	Feedback ($n = 103$ couples)		TAU ($n = 102$ couples)	
Outcome classification	*N*	%	*N*	%
Deteriorated	2	1.9	4	3.9
No change	14	13.6	24	23.5
Reliable change	10	9.7	12	11.8
Clinical significant change	42	40.8	11	10.8
At-risk couples[a]	56	54.4	76	74.5
At-risk responded	16	28.6	7	9.2
	Follow-up feedback ($n = 42$ couples)		Follow-up TAU ($n = 32$ couples)	
Deteriorated	0	0	1	3.1
No change	3	7.1	10	31.3
Reliable change	5	11.9	2	6.3
Clinically significant change	15	35.7	4	12.5

Note. ORS = Outcome Rating Scale (Miller & Duncan, 2004); TAU = treatment as usual.
[a] Either one or both in couple were at risk compared to expected treatment responses at Session 3.

Table 4
Multilevel Models Predicting Follow-Up on Outcome Rating Scale

Effects	Model 1 coefficient	Model 2 coefficient	Model 3 coefficient
Fixed effects			
Intercept (γ_{00})	26.69***	24.44***	24.43***
Client ORS (π_{10})	0.36***	0.38***	0.38***
Feedback (β_{01})		3.97**	3.98*
Random effects			
Couple variance (σ^2_{couple})	19.70***	15.63***	15.51**
Therapist variance (σ_{ther})	.009	.004	.003
Client variance (σ^2_e)	44.19	44.34	44.36
Slope of feedback (U_{02})			0.104
ρ_{couple}	0.31	0.26	
ρ_{ther}	0.00	0.00	

Note. ORS = Outcome Rating Scale (Miller & Duncan, 2004).
* $p < .05$. ** $p < .01$. *** $p < .001$.

pretreatment LW was 87.65, indicating that LW scores improved from pretreatment to follow-up ($\gamma_{00} = 87.65$, $p < .001$). The standardized mean difference of pretreatment and follow-up LW was $d = 0.36$. The intact couples in the feedback condition were more satisfied (accounting for pretest satisfaction) than the intact couples in the TAU condition, although the coefficient for feedback was not significant in the multilevel model ($\beta_{01} = 7.29$, $p = .124$). However, the size of the feedback effect was quite large ($d = 0.30$); the lack of significance appears to be due to the reduced power, given the lower sample size of intact couples.

Discussion

The present study tested the effects of feedback in couple therapy compared with couples receiving TAU in a naturalistic setting. Consistent with our hypothesis, the feedback condition emerged as significantly superior to TAU. A moderate to large ES (0.50) was found for the feedback condition. The predicted score adjusted for severity of an average client in the feedback group was 4.89 points higher than an average client in the TAU group. The difference was, in effect, the difference required for reliable change. Said another way, the average posttreatment score for persons in the feedback condition (26.35) was nearly 5 points greater than the average post score for those in the TAU group (21.69). The difference between the groups, in other words, nearly constituted both a reliable change and transcended the clinical cut off. The strong effect of feedback seems particularly noteworthy, given the relative simplicity of the intervention and in light of the comparison group being an active treatment.

In addition, the significant superiority of feedback over TAU was maintained at follow-up, although diminished, on the general outcome measure (ORS) but not on the one specific to relationship adjustment (LW). The continued advantage of feedback occurred on the primary measure even though the comparison was arguably biased toward TAU, given the differential numbers of separated or divorced couples who were not included in the follow-up analyses. Had separated couples been included, it is possible that the stronger effect on the ORS would have remained and an effect would have been found on the LW. Supporting this possibility is the finding that the TAU group had a 34.2% rate of separation or divorce versus 18.4% for the feedback condition.

In the feedback group, 40.8% (both in couple) scored 25 or above and 5 or more points in change compared with 10.8% of nonfeedback couples, a nearly fourfold difference. Six-month follow-up revealed that feedback maintained nearly a threefold advantage in proportion of clinically significant change (35.7% vs. 12.5%). The proportion of clinically significant change (40.8%) corresponds to Shadish and Baldwin's (2003) calculation of 40%–50% success rate in RCTs and suggests that similar rates can be achieved in clinical settings; however, it is less than the 48% reported in a recent large-scale RCT of couples (Christensen et al., 2004). The Christensen et al. (2004) study, however, averaged 22.9 sessions (vs. 4.8 sessions for the intervention in the current study), and the therapists were extensively trained and supervised.

The findings of the current study support continued reflection about the transportability of specific couple therapy approaches to clinical settings. As noted, couple therapy research has robustly demonstrated superiority over no-treatment controls for several approaches but has failed to find reliable superiority of one over another or TAU, especially at follow-up. At the same time, the financial investment for agency-wide implementation of a particular couple therapy orientation is substantial. For example, certification in emotionally focused couple therapy requires a minimum of 42 hr training and 32 hr of supervision with a certified EFCT supervisor (see http://www.eft.ca/training2.htm). This time and cost investment, in the context of high turnover in agencies, challenges the financial practicality of implementing approaches that have demonstrated efficacy only over no treatment (Sparks & Duncan, in press).

Conversely, the feedback condition in the current study demonstrated superior results to TAU at posttreatment and follow-up. The methods are generic in nature and not tied to a single therapy modality, and therefore represent a lower commitment of staff and money to implement. Therapists in the current study received only 17 hr of training. Feedback, therefore, seems more easily transportable to community settings, compared with specific treatment packages, and more likely to yield a return on investment.

The findings of the current investigation are consistent with the effects reported in the Reese et al. (2008) and Miller et al. (2006) studies, which used the same measures. Similar to the Lambert trials (see Lambert, in press), feedback with at-risk couples significantly improved outcomes over TAU couples at risk. Feedback, surprisingly, provided a preventive effect. Significantly fewer at-risk cases (those not proceeding according to ETR) emerged in the feedback condition.

The finding of an overall feedback effect for all clients is consistent with Harmon et al. (2007) and Hawkins et al. (2004). Hawkins et al. suggested that the provision of progress information to clients and therapists has more global effects than when feedback is provided only to therapists. More research is needed to investigate the impact of client involvement. PCOMS is a "client-directed" (Duncan et al., 2004) clinical process, and it is unknown how much therapist–client collaborative outcome and alliance monitoring impacts the feedback effect. It is also unknown how much the continuous alliance monitoring contributes to the feedback effect versus alliance assessment as a support tool for deteriorating clients (Whipple et al., 2003). Similarly, it is unknown how the supplemental use of alliance and other measures accounts

for an incremental increase in effectiveness. Any increase in client engagement with the therapist or between partners may have influenced or increased the feedback effect.

Therapist variance was found to be somewhat lower than reported in other studies, but there was significant variability in the effect of feedback across therapists. On the basis of only 10 therapists, a strong but preliminary negative relationship was found between therapist effectiveness without feedback and the size of the feedback effect; therapists at the lower end of effectiveness benefited more from feedback than their more successful colleagues. More research with a larger pool of therapists is needed to confirm this interesting finding. It is worth noting that 9 of 10 therapists did benefit from the effects of feedback.

There are several limitations to this study. First, the use of only one outcome measure in the pre–post treatment analysis limits the conclusions that may be drawn. In addition, the instruments used were quite brief, potentially limiting the understanding of the impact of feedback. We do not know if more extensive alliance and progress assessments would have given different results or whether other measures from clinician or observer perspectives would alter our findings. This study was intentionally pragmatic to more closely replicate what happens in routine clinical practice. The significant differences in the separation or divorce rates of the follow-up couples and the findings of the LW in the direction of the feedback group support the results found at posttreatment but do not eliminate the problems associated with the use of just one measure in the primary analysis.

Although the routine practice context of this study is noteworthy compared to settings where the sample is biased toward lower age and less severe presenting problems, it is unclear if the feedback effect found here can be shown in other couple therapy sites—for instance, in more ethnically diverse service contexts. As noted, the ETR trajectories were based on individual responses to treatment and therefore were not specific to couple therapy. This exerted an unknown effect on the feedback process. The data from the current study, however, will enable trajectories for couples to be determined.

The mechanisms of change that occurred in the feedback condition are not known. Providing outcome information to clients may have resulted in demand characteristics that favored the feedback condition (Orne, 1962). Clients may have been influenced to respond in a more socially desirable way when reinforced for apparent changes or when their lack of change was not fitting expected trajectories. We suspect, however, that the distress associated with relational problems would mitigate any tendency toward exaggerating improvement. Follow-up results support the effects of feedback and seem to diminish the likelihood of demand characteristics, given that clients completed the measures at home without the potential influence of the therapist.

The effects of feedback could have emerged from common factors (Duncan, Miller, Wampold, & Hubble, in press) or the increased attention to the monitoring of outcome and the alliance. The ritual of monitoring may have enhanced client expectancy, amplifying participation and securing a strong alliance. Alternately (or concurrently), clients may have been sensitized to the experience of change, thereby amplifying their efforts. The how and why of feedback warrants further exploration.

Allegiance effects are always a possible influence in any research (Luborsky et al., 1999)—both researcher effects and spillover allegiance effects to therapists. Although therapists in the study were familiar with the ORS/SRS system, they did not use the measures and believed that informal feedback from clients could suffice. In addition, therapists served as their own controls—the two treatment conditions did not draw from different therapist pools—so no special allegiances were promoted in one set of therapists in the experimental treatment versus another set of therapists in the TAU condition. Nevertheless, the principle investigator provided training for the therapists. An allegiance bias could have been transmitted through the researcher, influencing results in favor of the feedback condition.

To crudely address this possibility, at the end of the data collection phase project therapists again completed the attitude survey about attaining feedback and were asked an additional question: "Did the feedback cases turn out better than the TAU ones, did the TAU cases turn out better, or was there no difference?" Five therapists believed their feedback cases were better, 3 therapists felt that it did not matter, and 1 therapist believed TAU cases turned out better (1 therapist had left the agency). Moreover, the attitude survey showed a small decline in mean scores from prestudy to poststudy, suggesting that therapists, on average, did not develop more enthusiasm for feedback in the course of the study. Together, these results suggest that allegiance to feedback could account for some but not all of the advantage of the experimental condition.

The substantial benefits of feedback, supported by a growing empirical base that includes different measurement systems with varied populations in diverse contexts, suggests that the time has arrived for routine monitoring of outcomes and the use of continuous feedback to tailor and improve psychotherapy services. Two prominent groups within APA have recommended routine assessment of client response to treatment: the Division 29 Task Force on Empirically Supported Relationships (Ackerman et al., 2001) and the Presidential Task Force on Evidence-Based Practice (APA Presidential Task Force on Evidence-Based Practice, 2006; also ratified by the Norwegian Psychological Association, Norsk Psykologforening, 2007). Proponents from both sides of the common versus specific factors aisle have recognized that outcome is not guaranteed, regardless of evidentiary support of a given technique or the expertise of the therapist. A continuous feedback or practice-based evidence approach individualizes psychotherapy based on treatment response and client preference; systematic feedback addresses the dropout problem, as well as treatment and therapist variability, and could increase consumer confidence in the outcome of therapeutic services.

References

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

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285.

Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology, 19,* 98–110.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance–outcome correlation: Exploring the relative importance of ther-

apist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75,* 842–852.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., et al. (2001). Service profiling and outcome benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69,* 184–196.

Baucom, D., & Epstein, N. (1990). *Cognitive–behavioral marital therapy.* New York: Brunner/Mazel.

Benish, S., Imel, Z. E., & Wampold, B. E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review, 28,* 746–759.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., et al. (2004). Therapist effects. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 227–306). New York: Wiley.

Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72,* 176–191.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Derogatis, L. R. (1992). *SCL-90-R. Administration, scoring and procedure manual.* Baltimore: Clinical Psychometric Research.

Duncan, B. L., Miller, S. D., Reynolds, L., Sparks, J., Claud, D., Brown, J., & Johnson, L. D. (2003). The session rating scale: Psychometric properties of a "working" alliance scale. *Journal of Brief Therapy, 3*(1), 3–12.

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* (revised ed.). San Francisco: Jossey-Bass.

Duncan, B. L., Miller, S. D., Wampold, B. E., & Hubble, M. A. (in press). *The heart and soul of change: Delivering what works* (2nd ed.). Washington, DC: American Psychological Association.

Dunn, R. L., & Schwebel, A. I. (1995). Meta-analytic review of marital therapy outcome research. *Journal of Family Psychology, 9,* 58–68.

Freeston, M. H., & Plechaty, M. (1997). Reconsideration of the Locke-Wallace Marital Adjustment Test: Is it still relevant for the 1990s? *Psychological Reports, 81,* 419–434.

Goldman, A., & Greenberg, L. (1992). Comparison of integrated systemic and emotionally focused approaches to couple therapy. *Journal of Consulting and Clinical Psychology, 61,* 6–15.

Gollan, K. K., & Jacobson, N. S. (2002). Developments in couple therapy research. In H. A. Liddle, D. A. Santisteban, R. F. Levant, & J. H. Bray (Eds.), *Family psychology: Science-based interventions* (pp. 105–122). Washington, DC: American Psychological Association.

Greenberg, L. S., & Johnson, S. M. (1988). *Emotionally focused couples therapy.* New York: Guilford Press.

Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology, 58,* 1157–1172.

Hahlweg, K., & Klann, N. (1997). The effectiveness of marital counseling in Germany: A contribution to health services research. *Journal of Family Psychology, 11,* 410–421.

Hansen, N. B., Lambert, M. J., & Forman, E. V. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9,* 329–343.

Harmon, S. J., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, 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.

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.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51,* 1059–1064.

Jacobson, N. S., Follette, W. C., Revenstorf, D., Baucom, D. H., Halweg, K., & Margolin, G. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. *Journal of Clinical and Consulting Psychology, 52,* 497–504.

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.

Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin, 119,* 138–148.

Lambert, M. J. (in press). 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.* Washington, DC: American Psychological Association.

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

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* (5th ed., pp. 139–193). New York: Wiley.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians routinely to track patient outcome? A meta-analysis. *Clinical Psychology, 10,* 288–301.

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 patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11,* 49–68.

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

Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living, 21,* 251–255.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6,* 95–106.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Counseling and Clinical Psychology, 68,* 438–450.

Miller, S. D., & Duncan, B. L. (2004). *The Outcome and Session Rating Scales: Administration and scoring manual.* Fort Lauderdale, FL: Author.

Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, B. (2006). Using outcome to inform and improve treatment outcomes. *Journal of Brief Therapy, 5,* 5–22.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J., & 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., Sorrell, R., & Brown, G. S. (2005). The Partners for Change Outcome Management System. *Journal of Clinical Psychology, 61,* 199–208.

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A

preliminary study. *Journal of Consulting and Clinical Psychology, 76,* 116–124.

Norsk Psykologforening [Norwegian Psychological Association]. (2007). Prinsipperklæring om evidensbasert psykologisk praksis [Declaration on evidence-based psychological practice]. *Tidsskrift for Norsk Psykologforening, 9,* 1127–1128.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62,* 1157–1172.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Penn, Schoen, & Berland Associates (2004). [Survey for the American Psychological Association]. Unpublished survey.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2005). *HLM6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Reese, R. J., Norsworthy, L., & Rowlands, S. (2008). *Does a continuous feedback model improve psychotherapy outcomes?* Manuscript submitted for publication.

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy, 29,* 547–570.

Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology, 73,* 6–14.

Shadish, W. R., Montgomery, L., Wilson, P., Wilson, M., Bright, I., & Okwumabua, T. (1993). The effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61,* 992–1002.

Snyder, D. K., & Wills, R. M. (1989). Behavioral versus insight-oriented marital therapy: Effect on individual and interspousal functioning. *Journal of Consulting and Clinical Psychology, 61,* 61–69.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28.

Sparks, J. A., & Duncan, B. L. (in press). Couple and family therapy and the common factors: Have all won prizes? In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works* (2nd ed.). Washington, DC: American Psychological Association.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Hillsdale, NJ: Erlbaum.

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50,* 59–68.

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24,* 190–195.

Received July 31, 2008
Revision received March 26, 2009
Accepted April 3, 2009 ■