Published in final edited form as:

Psychother Res. 2009 March ; 19(2): 181–193. doi:10.1080/10503300802688460.

Correspondence of Motivational Enhancement Treatment Integrity Ratings among Therapists, Supervisors, and Observers

Steve Martino¹, Samuel Ball¹, Charla Nich², Tami L. Frankforter², and Kathleen M. Carroll²

¹Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA.

²Division of Substance Abuse, Yale University School of Medicine, New Haven, Connecticut, USA.

Abstract

This study examined the correspondence of treatment integrity ratings (adherence and competence) among community program therapists, supervisors, and observers for therapists who used motivational enhancement therapy (MET) within a National Institute on Drug Abuse Clinical Trial Network protocol. The results suggested there was reasonable agreement between the three groups of raters about the presence or absence of several fundamental MET strategies. Moreover, relative to observers, therapists and supervisors were more positive in their evaluations of the therapists' MET adherence and competence. These findings underscore the need for objective monitoring of therapists' performance when using empirically supported treatments and for adequately training therapists and supervisors to evaluate their treatment implementation in community programs, and are consistent with observations that different perspectives on the therapeutic process are not interchangeable.

Keywords

process research; psychotherapist training; substance abuse; motivational enhancement therapy; treatment integrity

Introduction

The virtual requirement that community programs implement empirically supported psychotherapy treatment has led to a related need to assess the integrity of treatment implementation by therapists who work in these settings (APA Presidential Task Force, 2005; Carroll & Rounsaville, 2007; Miller, Zweben, & Johnsen, 2005). Treatment integrity refers to how much therapists have implemented psychotherapies consistent with the principles and techniques of the targeted approach (i.e., adherence) and the skill with which therapists have delivered them (i.e., competence) (Waltz, Addis, Koerner, & Jacobson, 1993). If community program therapists are to achieve the improved treatment outcomes that have been the promise of using empirically supported psychotherapies, the integrity with which they implement those treatments and the manner in which integrity is assessed become critical issues (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). This study examines the correspondence of treatment integrity ratings provided by therapists, their supervisors, and observers in a multisite clinical trial of motivational enhancement therapy (Ball et al., 2007).

Three methods within psychotherapy clinical trial studies are commonly used to assess treatment integrity. These methods tap the perspectives of therapists, supervisors, and observers, respectively. Another method, having clients provide their perception of their

therapists' interventions in sessions (Elliot & Williams, 2003; Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; McCarthy & Barber, in press; Silove, Parker, & Manicavasgar, 1990), is seldom used in clinical trials for treatment integrity assessment and not the focus of this paper.

The first method involves the use of therapist checklists. In this approach, therapists complete brief checklists at the end of their sessions to indicate which interventions they delivered to clients consistent with the manual guided treatments (see Carroll, Nich, & Rounsaville, 1998). Therapist checklists take relatively little time to complete and aim to (a) help therapists learn new skills by monitoring their own performances, (b) enhance the quality of supervision by having therapists and supervisors focus on a common set of therapeutic strategies, and (c) establish the extent to which the treatment is actually used in clinical practice from the therapists' perspective. Some studies have shown that therapists' adherence ratings positively predict their clients' primary treatment outcomes (Henggeler et al., 1997; Henggeler, Pickrel, & Brondino, 1999), though this relationship has seldom been evaluated in the literature.

A second method for assessing treatment integrity is to have supervisors, who are proficient in the targeted treatment and trained in how to rate tape recorded sessions, evaluate those sessions for adherence and competence. During the trial, supervisors provide therapists with performance feedback and coach them to maintain and further develop their treatment skills (Baer et al., 2007; Bellg et al., 2004; Carroll, 1997). This approach has the advantage of relying on recorded data of what actually occurred in sessions rather than relying solely on therapist memory (Hill & Lambert, 2004). To our knowledge, the relationship between supervisors' ratings of therapists' treatment integrity and client treatment outcomes has not been examined in prior studies.

Third, many efficacy trials include ratings by observers who assess the therapists' treatment integrity from recorded sessions using psychotherapy process rating measures to establish that the targeted treatment approaches have been implemented with sufficient adherence and competence and to determine if different treatments in the trial are discriminable from each other (Perepletchikova & Kazdin, 2005). Observers are viewed as relatively independent in that they do not participate in the trials, are not informed about the treatment conditions or sessions to be evaluated, and are not aware of the hypotheses being tested. Given the extensive training observers typically receive in such trials, their ratings usually are highly reliable, especially when observers tally every occurrence of each rating item (Weiss, Marmar, & Horowitz, 1988). Moreover, in many studies observers' ratings of adherence and competence (Barber, Crits-Christoph, & Luborsky, 1996; DeRubeis & Feeley, 1990; Hoffart, Sexton, Nordahl, & Stiles, 2005; Miller, Benefield, & Tonigan, 1993; O'Malley et al., 1988; Shaw, Elkin, Yamaguchi, Olmstead, & Vallis, 1999) have been found to relate positively with client treatment outcomes. However, these relationships often are very complex (e.g., other variables such as alliance may moderate the effect of adherence on outcome) and findings have been inconsistent across studies (Barber et al., 2006; Barber, Sharpless, Klostermann, & McCarthy, 2007; Beutler et al., 2004).

Little data exist that examine the extent to which treatment integrity ratings derived by these three methods correspond with each other. Carroll et al. (1998) reported on the concordance of therapists' and independent observers' assessments of the therapists' use of specific cognitive behavioral therapy (CBT) interventions for cocaine abuse. They found poor overall therapist-observer agreement on dichotomous assessments (present, absent) of therapists' use of CBT interventions (e.g., 80% of the item kappas were in the poor to fair range). Therapists in that trial reported significantly greater use of specific CBT techniques relative to observers. More recently, two studies by Miller and colleagues (2001; 2004) compared post-workshop global proficiency ratings made by therapists trained in motivational interviewing with several

proficiency standards derived from independent observers' evaluation of therapists' actual motivational interviewing sessions. Both studies found that the therapists' self-reported higher overall proficiency in motivational interviewing than determined by the observers. Similarly, there is very little data examining how supervisors evaluate therapists' treatment integrity. Borders and Fong (1991) found a weak and insignificant relationship ($r = .12$) between supervisors' global ratings of trainees' counseling skills and ratings of counseling competencies as determined by two trained observers evaluating audiotapes of the same sessions. No other studies have directly compared how supervisors evaluate treatment integrity relative to observers, and none have compared supervisor and therapist evaluations of the same sessions.

Treatment integrity ratings by therapists, supervisors, and observers are all subject to bias, as each has unique perceptions about the sessions being evaluated, which might affect their ratings and contribute to the low correspondence among rater groups. For example, therapists directly experience what happens in sessions, giving them privileged information (e.g., cues and nonverbal communication, contextual factors influencing treatment) about their clients that might not be evident in recordings, yet may influence their self-assessments of treatment integrity (Hill & Lambert, 2004). Therapists also may overestimate the quantity and quality of their interventions when they know they are being assessed (Perepletchikova & Kazdin, 2005) or evaluate sessions more positively when self-assessing their performance immediately after sessions rather than when evaluating their sessions in tape-assisted reviews (Hill et al., 1994). Supervisors typically monitor the work of several therapists over time, giving them information about the relative performance of therapists. Depending on their prior assessments of these therapists, they might be prone to be more lenient or severe in their judgments of therapists' performance in subsequent sessions (Hoyt, 2000). Supervisors also may experience demand characteristics to evaluate the therapists' treatment implementation favorably when they want to show how therapists have benefited from their supervision. Independent observers also may experience biases in that they may interpret specific rating items idiosyncratically (Hill & Lambert, 2004), rate therapists they perceive as similar to themselves positively (Hill, O'Grady, & Price, 1988; Mahalik, Hill, O'Grady, & Thompson, 1993), or rate more leniently or severely across items due to general impressions not grounded in rating item definitions (Hoyt, 2000). On the other hand, they typically are carefully trained to interpret the meaning of rating items similarly, a process intended to encourage greater objectivity and reduce subjective perception when rating therapists' treatment performance (Hill & Lambert 2004; Hoyt, 2000).

The extent to which therapists, supervisors, and observers agree about treatment integrity using these different methods could inform how community programs establish that therapists have implemented treatments consistent with recommended practice guidelines. If therapists consistently evaluate their treatment adherence and competence similarly to supervisors and observers, then the most simple, least costly approach for measuring treatment integrity would be to use therapist checklists. Supervisory assessment requires an additional burden of training supervisors how to rate treatment integrity and to use the ratings to further develop therapists' treatment skills. However, onsite supervisors have the advantage of providing therapists with ongoing treatment integrity monitoring and training over time (Carroll et al., 2002). Alternatively, observers often have more rating expertise than therapists and supervisors. If available to community programs, observers could provide periodic treatment integrity assessments to programs without shifting this demand to therapists or supervisors. This model has been used to foster treatment integrity when complex psychotherapies such as multisystemic therapy (Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002), dialectical behavioral therapy (Dimeff & Koerner, 2007), and integrated dual diagnosis assertive community treatment (McHugo, Drake, Teague, & Xie, 1999) have been implemented in large systems.

In this report we describe the correspondence of treatment integrity ratings provided by therapists, their supervisors, and independent observers who participated in a multisite National Institute on Drug Abuse Clinical Trial Network effectiveness protocol (Ball et al., 2007) evaluating motivational enhancement therapy (MET), a manualized version of motivational interviewing (Miller & Rollnick, 2002) originally developed for Project MATCH (Miller, Zweben, DiClemente, & Rychtarik, 1992). MET is a brief intervention in which therapists use empathic counseling techniques (such as reflective listening) and strategies for eliciting client self-motivational statements (such as using feedback to produce discomfort with status quo behaviors) to enhance a clients' intrinsic motivation for behavioral change. MET has demonstrated moderate treatment effects (.4 to .5) for substance use disorders (Burke, Arkowitz, & Menchola, 2003; Hettema, Steele, & Miller, 2005). We hypothesized that, consistent with prior reports, therapists' ratings of their own adherence to MET would be higher than observer ratings. Second, we hypothesized that supervisors trained to rate the sessions using an adherence and competence scale would have higher rates of agreement with the observers for both MET adherence and competence ratings than with the therapists' ratings.

Method

Overview of MET protocol

The MET protocol was implemented within five outpatient community treatment programs that served diverse samples of substance users. Programs were located in California, Connecticut, and Pennsylvania in suburban and urban settings. Clients seeking outpatient substance abuse treatment at each site were randomized to receive either three individual MET or counseling-as-usual (CAU) sessions. Program therapists conducted the sessions during the first four weeks of treatment. Therapists who delivered MET also received supervision from program supervisors who rated the therapists' MET adherence and competence and provided them with feedback and coaching to support their implementation of MET. Observers subsequently independently evaluated the adherence and competence of MET and CAU sessions to establish MET treatment integrity in the trial. The primary findings from the MET protocol suggested that both MET and CAU reduced substance use at the end of the 4-week treatment phase; however, during the 12-week follow-up period, participants assigned to MET sustained reduced primary substance use, whereas those in CAU significantly increased use over this same phase. In addition, MET therapists used MET strategies significantly more often and with greater competence than CAU therapists (see Ball et al., 2007 and Martino, Ball, Nich, Frankforter, & Carroll, 2008 for detailed description of the study and its findings).

Participants

Participants were the therapists and supervisors at the five study sites involved in the delivery of MET, the observers who evaluated the therapists' treatment integrity, and the clients who received MET treatment.

Therapists—Volunteers were drawn from the staff of the participating treatment programs based on their willingness to be randomized to deliver either MET or CAU and to have their sessions audiotaped. The 14 therapists who were trained in and implemented MET within the protocol were participants in this study. Ratings from the therapists who were randomized to deliver CAU were not included in this study because they did not rate their session performance or participate in rating-based supervision. Therapists provided either written permission or informed consent for participation depending on local Institutional Review Board requirements. Most MET therapists had no prior motivational interviewing or MET training exposure, and none had been trainers or therapists in research studies involving MET (Ball et al., 2002a). Therapists were predominantly female (60%) and Caucasian (77%). On average, therapists were 38.9 years old ($sd = 11.8$), employed at their agencies for a mean of 3.2 years

($sd = 3.9$), had been working as substance abuse counselors for 8.1 years ($sd = 6.4$), and completed 14.5 years of education ($sd = 5.1$). Forty-three percent had a master's degree, 46% were state certified substance abuse counselors, and 45% indicated they were in recovery from prior substance abuse problems.

Supervisors—Four supervisors evaluated MET sessions across five sites.¹ Protocol supervisors were only involved in the monitoring of the MET intervention and had no equivalent involvement in the CAU condition. Supervisors provided either written permission or informed consent for participation depending on local Institutional Review Board requirements. Three of the four supervisors were female, all Caucasian, on average 40.2 years old ($sd = 7.7$), and had a master's degree or higher. On average, they had been providing substance abuse treatment services for 12.0 years ($sd = 6.8$), supervision for 8.2 years ($sd = 5.6$), and had 13.0 hours ($sd = 8.4$) of formal motivational interviewing or MET training prior to protocol involvement.

Observers—Fifteen observers evaluated the session audiotapes. On average, observers were 37.7 years old ($sd = 9.7$) and female (53%). Most observers had master's degrees in a clinical profession (67%) and had an average of 6.9 years ($sd = 9.7$) of substance abuse treatment experience, 8.3 years ($sd = 7.9$) of general psychotherapy experience, and 5.6 years ($sd = 5.3$) of clinical research experience. Sixty percent of the observers had served as independent raters in prior clinical trial studies testing the efficacy of behavioral treatments. Fifty-three percent reported prior workshop training in motivational interviewing or MET for an average of 9.0 hours ($sd = 5.9$).

Clients—Sixty-four clients who received treatment from the 14 MET therapists were participants in this study, representing 36% of the all the clients who received MET in the original study. Clients were English-speaking, with a range of primary substance use problems (38% alcohol, 16% marijuana, 14% cocaine, 11% opiates, 6% methamphetamines, 5% alcohol and drug, and 10% other). On average they were 33.4 years old ($sd = 10.6$), male (73%), primarily Caucasian (44%) or African American (38%), and single (84%). Clients completed on average 12.3 years ($sd = 1.8$) of education. Less than half of the clients were employed (35.9%). These characteristics were similar to the full sample of clients in the MET condition and the protocol overall.

Measures

Independent Tape Rater Scale (ITRS)—The ITRS (Martino et al., 2008) is a 39-item scale adapted from the Yale Adherence and Competence Scale (Carroll et al., 2000) to assess community program therapists' adherence and competence in implementing MET, strategies considered inconsistent with MET (e.g., direct confrontation) or common to drug counseling (e.g., assessing substance use), and several general characteristics of the therapists (e.g., ability to maintain the session's structure) and clients (e.g., initiation of discussions unrelated to the session). This study focused on the 10 items that were common across the scales used by the therapists, supervisors, and observers and that involved MET-specific interventions.

For each item, observers evaluated the therapists on two dimensions using a 7-point Likert scale. First, they rated the extent to which the therapist delivered the intervention (adherence; 1 = not at all, to 7 = extensively). Second, they rated the skill with which the therapist delivered the intervention (competence; 1 = very poor, to 7 = excellent). ICC estimates suggested excellent levels of interrater reliability for the 10 items for both the adherence (mean ICC = .

¹Two sites in the three-session protocol had the same person conduct MET supervision in the study. This supervisors' position and location permitted her to transverse both agencies.

89; range = .66 to .99) and competence dimensions (mean ICC = .85; range = .69 to .97). A confirmatory factor analysis supported a two-factor model for the 10 MET consistent items (Martino et al., 2008). Five items included MET fundamental strategies that underpin the empathic and collaborative stance of MET such as use of open-ended questions, reflective statements, and motivational interviewing style or spirit. The other five MET consistent items involved advanced or structured strategies for evoking client motivation for behavior change, such as heightening discrepancies and change planning. The fundamental and advanced MET strategies factors retained excellent inter-rater reliability, consistent with their individual components (adherence ICC: fundamental = .91; advanced = .95; competence ICC: fundamental = .89; advanced = .89). Table 1 describes the rating items that comprise each factor. A detailed description of the psychometric analysis of the ITRS is provided in another report (Martino et al., 2008).

Therapist Session Checklist—The Therapist Session Checklist asks therapists to indicate the frequency of their use of the ten MET consistent strategies in the ITRS. Therapists rated the extent to which they believed they had used the strategies during the session along 7-point Likert scales (1 = not at all, to 7 = extensively). Because therapists only rated their own sessions, interrater reliability was not determined. A confirmatory factor analysis using structural models with AMOS (6.0) software (Arbuckle, 2005) with maximum likelihood estimation showed the two-factor model demonstrated a good fit to the data, namely, a Root Square Mean Square Error of Approximation (RMSEA) of .05 or less (Browne & Cudeck, 1993) and the Normed Fit Index (NFI), Incremental Fit Index (IFI), and Comparative Fit Index (CFI) of .90 or greater (Hu and Bentler, 1995; Kline, 1998; Marsh, Balla, & McDonald, 1988; Yadama & Pandey, 1995). The respective fundamental and advanced strategies fit indices were as follows: RMSEA = .06 and .00; NFI = .98 and 1.0; IFI = .99 and 1.0; and CFI = .99 and 1.0.

Supervisor Tape Rater Scale—The supervisors rated taped sessions using a 30-item supervisory version of the ITRS that excluded rating items involving general characteristics of the therapists and clients. Supervisors rated both adherence and competence using the 7-point Likert scales as described for the ITRS. Interrater reliability was not determined because no more than one supervisor rated each session. Confirmatory factor analysis of the MET consistent adherence items as described above showed the two factor model fit the data adequately for the respective fundamental and advanced strategies (RMSEA = .03 and .00, NFI = .96 and .99; IFI = .99 and 1.0, and CFI = .99 and 1.0).²

Procedures

Therapist and Supervisor Training—All MET therapists and supervisors were trained by MET experts. The MET experts were master's or doctoral degreed clinicians who had extensive experience training and supervising clinicians in motivational interviewing and MET, had completed a 3-day MET trainer's workshop, and were trained in the use of the ITRS as part of the MET protocol. The MET experts conducted two-day intensive MET workshops for the therapists and supervisors at their sites, followed by individually supervised practice cases until minimal protocol certification standards had been achieved in three sessions (i.e., at least half of the MET integrity scale items rated average or above in terms of adherence and competence). In addition, the MET experts reviewed with therapists how to complete the Therapist Session Checklist at the end of each session and supervisors how to use Supervisor Tape Rater Scale. No formal training criteria were used to establish the accuracy with which therapists or supervisors used these scales. After therapists were certified in MET, they began to treat randomized clients in the protocol and receive biweekly supervision from their

²For brevity, tables of fit indices for models of therapist MET consistent adherence are not included here. They are available upon request from the corresponding author.

supervisors who provided the therapists with MET adherence and competence rating-based feedback and coaching after reviewing audiotaped client sessions. These therapist and supervisor training procedures are commonly used in clinical trials (Baer et al., 2007) and have been shown to result in significant improvements in therapists' practice of empirically supported treatments (Miller et al., 2004; Sholomskas et al., 2005).

Observer Training—Observers were trained to assess therapist adherence and competence using the ITRS (Martino et al., 2008). All observers attended an initial 8-hour didactic seminar in which they reviewed a detailed rating manual (Ball, Martino, Corvino, Morgenstern, & Carroll, 2002b) that specified item definitions and rating decision guidelines and practiced rating the items in both limited therapist-client transactions and in a full protocol session. Following initial training, each observer completed ratings for an identical set of 10 calibration tapes randomly selected from the larger pool of protocol tapes to include the two treatment conditions as well as different sites and sessions. Initial item reliabilities were calculated using Shrout and Fleiss (1979) ICC two-way mixed model (3.1), with item ratings as the fixed effect and observers as the random effect. Next, a second 6-hour rater training was held to address items with lower initial reliability; the observers then completed a set of five additional tapes for final inter-rater reliability calculation. To sensitize observers to specific items in which their ratings might drift, observers were informed if any of their ratings were more than two scale points above or below 'expert' (SM and SB) consensus ratings on the same set of calibration tapes. Combining the didactic and calibration tape components, each rater completed about 44 hours of training. To ensure ongoing rater reliability, a randomly selected common tape was rated on five separate occasions approximately four months apart. Observers were aware of this procedure, but not its timing. Once again, individual item ratings were compared to expert ones and feedback was provided to all observers.

Sessions—On average, MET sessions were 45.4 minutes long ($sd = 7.8$) and all of them were audiotaped. Data used for this study were taken from all MET sessions in which therapists, supervisors, and observers had rated the same session. Of the 451 total MET sessions in the protocol, 117 (26%) were completed across raters. This sample included all 14 therapists who delivered MET in the trial and was balanced across sessions (48 first sessions, 37 second sessions, and 32 third sessions). The reduction from the total number of sessions occurred because the supervisors did not rate all MET sessions, and the observers only rated a subsample of MET sessions.

Ratings by Therapists—At the end of each session, therapists completed the Therapist Session Checklist to indicate their self-assessment of their adherence to MET-consistent strategies. They did not rate their own competence in delivering these strategies because the higher level of inference involved in making competency determinations would have required additional rating training to reduce rater bias (Hoyt, 2000; Hoyt & Kerns, 1999).

Ratings by Supervisors—Supervisors rated an entire audiotaped session once per month, which was randomly selected by research staff for each MET therapist. These ratings formed the basis of MET performance feedback and coaching interventions to enhance the therapists' MET skills. To maintain consistency across supervisors, tapes were co-rated with MET experts on a monthly basis to promote accurate rating, supervisory feedback, and coaching of MET.

Ratings by Observers—Independent observers, blind to treatment condition, session number, and site, rated entire sessions within a substantial subset ($n = 425$, 44%) of the protocol tapes. Given the high reliability of the rating scale and checks for rating drift during the rating period, only one observer rated each taped session.

Results

Therapist, supervisor, and observer correspondence on adherence ratings

Correspondence of adherence ratings was evaluated in terms of 1) whether an intervention occurred (e.g., categorically present, regardless of how much it occurred in a given session) and 2) the extent to which it occurred (1 – 7 rating) by rater category (therapist, supervisor, observer). Table 2 shows the percentage of sessions in which MET strategies occurred at least once in the session according to the therapists, supervisors, and observers, including the percentage of absolute agreement among these raters. In general, there was good to very good agreement for 4 of 5 items tapping fundamental MET strategies across protocols. That is, absolute agreement was above 90% for the items regarding use of open-ended questions, reflective statements, and use of an overall motivational interviewing style and 74% for the item assessing use of affirmations. However, for all items evaluating more strategic applications of MET (i.e., all advanced strategies for eliciting client self-motivational statements), both therapists and supervisors estimated these strategies were present in more of the sessions than the observers (all absolute agreements < 65%).

For the continuous measure, we approached the analyses in two ways. First, we estimated the degree of consistency among the three groups of raters for the fundamental and advanced MET mean adherence ratings using Shrout and Fleiss (1979) intraclass correlation coefficients (ICC) two-way mixed model (3,1), with ratings as the fixed effect and raters as the random effect. We similarly calculated ICC estimates for the correspondence of fundamental and advanced MET mean competence ratings between supervisors and observers (note the therapists were not asked to rate their own competence levels). ICC estimates for the adherence ratings (e.g., agreement on how much the intervention occurred in each session) for fundamental and advanced MET strategies indicated there was poor agreement among the raters about how often on average fundamental and advanced MET strategies occurred (ICC = .34 and .33, respectively).

Second, we used generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Hoyt & Melby, 1999; Shavelson, Webb, & Rowley, 1989) to simultaneously examine the main effects of rater type (therapist, supervisor, observer), item type (fundamental, advanced) nested within rater type, and taped sessions evaluated (session 1, 2, or 3), as well as the interaction among these factors (called facets in generalizability theory) on the adherence and competence ratings. In generalizability theory, a generalizability coefficient denoted by G and analogous to an ICC in a single facet design (Mushquash & O'Connor, 2006), is calculated and can be used to make absolute (pass/fail) or relative decisions (e.g., examining differences between groups or individuals in MET proficiency) about the optimal level of each facet to reduce error variance and maximize reliability in measurements (Shavelson et al., 1989). For this study, relative G is most appropriate for examining the degree of consistency among the rater groups. Variance estimation and G-coefficient calculation was conducted using an SPSS program by Mushquash and O'Connor (2006).

Table 3 provides variance generalizability theory component estimates based on a two facet nested design to assess variability in adherence rating associated with rater type, item type, and session number. The generalizability coefficient was .60, suggesting low levels of rating reliability. The greatest amount of observed variation in adherence rating was explained by differences among the raters (39%) rather than by differences in the type of items rated (14%) or specific sessions taped (9%). Thirty-eight percent of the total variance was attributable to residual error.

Finally, we conducted one-way ANOVAs using a Bonferroni-corrected α of .025 (.05/02) to test for the two hypothesized rater differences in mean fundamental and advanced MET

adherence and competence ratings (see Table 4). One-way ANOVAs on therapist fundamental, $F(2,349) = 30.3, p < .001$, and advanced MET adherence scores, $F(2,349) = 127.4, p < .001$, indicated significant differences among the three sets of ratings. *Post hoc* Tukey tests showed that while therapists and supervisors estimated similar mean amounts of fundamental MET strategies occurring in sessions, their estimates were significantly higher than those provided by the observers. In addition, their estimates of advanced MET strategy frequency were significantly higher than observer estimates, though the estimates provided by the therapists were higher than those provided by the supervisors. Re-analysis of the data using only one randomly selected rated session for each client, to prevent the potential confounds of non-independence of observations, did not change these results.³

Supervisor and observer correspondence on competence ratings

ICC estimates for fundamental (ICC = .26) and advanced mean competence ratings (ICC = .29) showed poor agreement between supervisors and observers. Likewise, the G-coefficient (.63) and variance estimate components indicate low levels of rating reliability. As with adherence ratings, the greatest amount of observed variation was explained by differences among the raters (47%) rather than by differences in the type of items rated (13%) or specific sessions taped (12%), with 28% of the variance attributable to residual error (see Table 3). One-way ANOVAs suggested that supervisors rated the therapists' competence in fundamental MET delivery higher than how observers evaluated the therapists' performance in this area ($F(1,233) = 14.6, p < .001$), whereas these raters did not differ significantly from each other in their evaluation of therapists' advanced MET strategies ($p = .30$). This pattern of results remained the same in a re-analysis of the data using only one randomly selected taped session per client.

Discussion

This study was the first to report on the correspondence of therapists, supervisors, and observers on ratings of therapist adherence and competence based on a large number of sessions drawn from a national multisite trial of MET conducted within community programs. There were three major findings. First, when therapists, supervisors, and observers evaluated sessions categorically for the presence or absence of specific MET interventions, there was good to very good agreement among them on several fundamental MET strategies that underpin the approach's empathic and supportive stance (open-ended questions, reflective statements, affirmations, motivational interviewing style). There was less agreement about whether therapists used advanced MET strategies commonly employed to elicit client statements in favor of change. Second, there was poor agreement regarding the extent to which the specific interventions occurred within sessions. Therapists and supervisors provided higher ratings of fundamental and advanced MET strategy adherence relative to the observers. Therapists also rated their advanced MET strategy adherence as higher than supervisors. Higher levels of variability in adherence ratings were attributable to differences between the groups of raters rather than systematic differences in how raters approached rating individual items or specific sessions taped. Third, there was poor agreement between supervisors and independent observers about the therapists' competence levels, with the supervisors generally providing higher estimates than the observers for fundamental MET strategies but few differences for competency ratings of advanced MET strategies. As with the adherence ratings, more variation in competence ratings was attributable to differences between the raters than due to differences in how raters responded to individual item or sessions.

³A table presenting the results of the re-analysis of the mean fundamental and advanced MET strategy adherence and competence ratings among therapists, supervisors, and observers involving only one randomly selected taped session per client ($n = 64$) is available from the corresponding author.

These results parallel findings from prior studies (Carroll et al., 1998; Miller et al., 2004; Miller & Mount, 2001) that indicate low rates of correspondence between therapists' self-reports of their adherence and competence in delivering specific therapies and those of observers not involved in the treatment. This study extends previous research in this area by using a detailed continuous rating system of specific interventions rather than global therapist impressions of their overall proficiency (Miller & Mount, 2001; Miller et al., 2004) or only the presence or absence of strategies used within sessions (Carroll et al., 1998). In addition, this study goes beyond traditional methods for determining inter-rater agreement (e.g., absolute agreement, ICCs) by using generalizability theory to establish that a substantial proportion of the variation in adherence and competence ratings was due to actual rater differences rather than other facets. Moreover, this study's findings are relevant to clinical practice and dissemination efforts, as the therapists estimated they had used MET strategies several more times per session than as rated by observers. This difference in perspective could lead to a situation in which some therapists believe they have sufficiently adhered to MET practices when their performance, as evaluated supervisors or observers, may not reach proficiency standards.

This study also is the first investigation to compare program supervisors' assessments of therapists' treatment adherence and competence to those of independent observers. Supervisors rated therapists as implementing fundamental and advanced MET strategies more often and with more competence for the fundamental strategies than did the observers. Differences in adherence ratings between supervisors and observers were smaller in magnitude than between the therapists and observers, however. In addition, there was more agreement between the supervisors and observers regarding their ratings of therapists' competence levels, particularly for the advanced MET strategies. While the rating training and support provided to supervisors may have developed their adherence and competence rating abilities, supervisors might benefit from additional training and practice in how to rate tapes more consistently (Hill & Lambert, 2004; Hoyt, 2000) to provide therapists with more reliable and valid assessments of their MET integrity.

This study's findings raised the question of which methods for evaluating adherence and competence should be used to assess therapist treatment integrity in clinical practice. Because therapists, supervisors, and observers typically agreed about whether several fundamental MET strategies occurred within sessions, the use of therapists' self-reports from therapeutic strategy checklists may be a reasonable and cost effective means for determining if therapists have used these basic counseling strategies (e.g., open-ended questions, reflections, affirmations). These determinations may require less inference and, therefore, be easier for therapists to make without more extensive rating training (Hoyt, 2000; Hoyt & Kerns, 1999). However, simply establishing the presence or absence of basic therapeutic techniques may be of limited value without more fine-tuned ratings of level of adherence to more complex strategies, which were found to be areas of poorer rater correspondence in this study. For example, having twice as many reflections than questions or substantially more MET adherent strategies relative to all interventions are considered markers of proficiency in motivational interviewing (Miller & Mount, 2001; Miller et al., 2004). In this study, the observers received the most training and monitoring to interpret the meaning/occurrence of the items similarly. Training included alerting observers to potential rating biases (e.g., leniency/severity, halo effects), carefully reviewing each item's definition and rating decision rules, practicing rating items across a range of MET sessions, therapists, clients, and sites, calibration of inter-rater reliability until high agreement was reached, periodically checking the observers' performance relative to MET experts, and providing them with feedback to reduce drift (Martino et al., 2008). Supervisors received some training and monthly monitoring in their use of the rating system, but their level of training did not approach the rigor, feedback, and time involved in preparing the observers. In contrast, the therapists simply reviewed how to complete the checklist based upon their understanding of MET strategies derived from the workshop training

and biweekly supervision, without additional integrity rating training. Thus, while all raters may have been subject to bias, the observers were trained most thoroughly to reduce variability in their interpretation of the rating items. Their ratings likely represented the least biased depiction of the therapists' MET implementation.

If one assumes that the observers' ratings are most reliable or least biased, and if both therapists and supervisors are prone to overestimate the presence, frequency, and competence of the therapists' implementation of empirically validated therapies such as MET, it is possible that therapists and supervisors are more likely to rate a therapists' performance as sufficiently proficient when it may not be. This potential scenario could accentuate the problem of therapists believing and reporting they are implementing empirically supported treatments without actually having changed their treatment-as-usual practices (Miller, 2007; Miller, Sorensen, Selzer, & Brigham, 2006; Santa Ana et al., in press). Hence, this study serves as a reminder to be wary of drawing conclusions regarding the extent to which specific treatments are used in clinical practice based solely on what therapists say happened and adds an additional cautionary note about the complimentary positive bias supervisors may have about therapists' adherence and competence levels.

This study has several limitations. Some of the rating discrepancies may have been due to the different rating methods used by the therapists, supervisors, and observers rather than systematic differences in what the three rater groups believed occurred in the sessions. For example, therapists may have evaluated their performance from taped sessions more conscientiously than when self-reporting on their performance immediately after sessions (Hill et al., 1994). Second, it is possible that the therapists and supervisors, in being more involved in the cases, may have been more sensitive to certain aspects of the treatment than the observers who may not have recognized some interventions based on their rigorous reading of the item definitions to maximize reliability. Third, although observers' ICC estimates were comparatively good, variability in inter-rater reliability across MET items might partially explain rating discrepancies among therapists, supervisors, and observers. Fourth, the use of only one rater per session for each rater group may have diminished the validity of the treatment integrity measures (Hill & Lambert, 2004). Fifth, the generalizability of the findings were limited given the small number of therapists and supervisors included in this study, the inclusion of only programs included in the Clinical Trials Network, and the study's focus on MET integrity instead of additional treatment approaches. Finally, this study did not examine the relationship of the different adherence and competence ratings provide by therapists, supervisors, and observers to treatment outcomes. Thus, which perspective matters the most for predicting how clients might benefit from MET remains an open question.

Nonetheless, this study demonstrated how different perspectives about treatment integrity are not interchangeable. Future research should include the clients' perspective about what they believe their therapists are doing in sessions. This direction would require the development of treatment integrity measures that gather multiple perspectives on strategies used in treatment (McCarthy & Barber, in press). In addition, the predictive validity of client, therapist, supervisor, and observer integrity ratings on treatment processes and outcomes needs to be established. It is possible that the perspectives from different rater groups might have unique relations to treatment processes and outcomes or that a shared or composite evaluation may be the best indicator of the therapists' treatment adherence and competence.

References

- American Psychological Association Presidential Task Force. Report of the 2005 presidential task force on evidence-based practice. 2005. Retrieved December 1, 2008, from <http://www.apa.org/practice/ebpreport.pdf>

- Arbuckle, J.L. AMOS (Version 6.0). SPSS; Chicago, IL: 2005. Computer Program
- Baer JS, Ball SA, Campbell BK, Miele GM, Schoener EP, Tracy K. Training and fidelity monitoring of behavioral interventions in multi-site addictions research: A review. *Drug and Alcohol Dependence* 2007;87:107–118. [PubMed: 17023123]
- Ball SA, Bachrach K, DeCarlo J, Farentinos C, Keen M, McSherry T, et al. Characteristics, beliefs, and practices of community therapists trained to provide manual-guided therapy for substance abusers. *Journal of Substance Abuse Treatment* 2002;23:309–318. [PubMed: 12495792]
- Ball SA, Martino S, Corvino J, Morganstern J, Carroll KM. Independent tape rater guide. 2002bUnpublished psychotherapy tape rating manual
- Ball SA, Martino S, Nich C, Frankforter TL, Van Horn D, Crits-Christoph P, et al. Site matters: Motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology* 2007;75:556–567. [PubMed: 17663610]
- Barber JP, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcomes in brief dynamic therapy. *Journal of Consulting and Clinical Psychology* 1996;64:619–622. [PubMed: 8698958]
- Barber JP, Gallop R, Crits-Christoph P, Frank A, Thase ME, Weiss RD, et al. The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research* 2006;16:299–240.
- Barber JP, Sharpless BA, Klostermann S, McCarthy KS. Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice* 2007;38:493–500.
- Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 2004;23:443–451. [PubMed: 15367063]
- Beutler, L.E.; Malik, M.; Alimohamed, S.; Harwood, T.M.; Talebi, H.; Noble, S., et al. Therapist variables. In: Lambert, M.J., editor. *Bergin and Garfield's handbook of psychotherapy and behavior change*. Vol. 5th Edition. John Wiley and Sons, Inc.; New York: 2004. p. 227-306.
- Borders L, Fong ML. Evaluations of supervisees: Brief commentary and research report. *Clinical Supervisor* 1991;9:43–51.
- Browne, M.W.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, K.A.; Long, J.S., editors. *Testing structural equation models*. Sage; Newbury Park, CA: 1993. p. 132-162.
- Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology* 2003;71:843–861. [PubMed: 14516234]
- Carroll KM. New methods of treatment efficacy research: Bridging clinical research and clinical practice. *Alcohol Health & Research World* 1997;21:352–358. [PubMed: 15706748]
- Carroll KM, Farentinos C, Ball SA, Crits-Christoph P, Libby B, Morgenstern J, et al. MET meets the real world: design issues and clinical strategies in the Clinical Trials Network. *Journal of Substance Abuse Treatment* 2002;23:73–80. [PubMed: 12220604]
- Carroll KM, Nich C, Rounsaville BJ. Utility of therapist session checklists to monitor the delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research* 1998;8:370–320.
- Carroll KM, Nich C, Sifry R, Frankforter T, Nuro KF, Ball SA, Fenton LR, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence* 2000;57:225–238. [PubMed: 10661673]
- Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. *Addiction* 2007;102:850–862. [PubMed: 17523974]
- Cronbach, L.J.; Gleser, G.C.; Nanda, A.N.; Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. John Wiley; New York: 1972.
- DeRubeis R, Feeley M. Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research* 1990;14:469–482.
- Dimeff, L.A.; Koerner, K. *Dialectical behavior therapy in clinical practice*. Guilford Press; New York: 2007.

- Elliot M, Williams D. The client experience of counselling and psychotherapy. *Counselling Psychology Review* 2003;18:34–38.
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. Implementation research: A synthesis of the literature. University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231); Tampa, FL: 2005.
- Henggeler SW, Melton GB, Brondino MJ, Scherer DG, Hanley JH. Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology* 1997;65:821–833. [PubMed: 9337501]
- Henggeler SW, Pickrel SG, Brondino MJ. Multisystemic treatment of substance abusing and dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research* 1999;1:171–184. [PubMed: 11258740]
- Henggeler SW, Schoenwald SK, Liao JG, Letourneau EJ, Edwards DL. Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Child Clinical Psychology* 2002;31:155–167.
- Hettema J, Steele J, Miller WR. Motivational interviewing. *Annual Review of Clinical Psychology* 2005;1:91–111.
- Hill, CE.; Lambert, MJ. Methodological issues in studying psychotherapy processes and outcomes. In: Lambert, MJ., editor. *Bergin and Garfield's handbook of psychotherapy and behavior change*. Vol. 5th edition. John Wiley and Sons, Inc.; New York: 2004. p. 84-135.
- Hill CE, O'Grady KE, Balenger V, Busse W, Falk DR, Hill M, et al. Methodological examination of videotape-assisted reviews in brief therapy: Helpfulness ratings, therapist intentions, client reactions, mood, and session evaluation. *Journal of Counseling Psychology* 1994;41:236–247.
- Hill CE, O'Grady, Price P. A method for investigating sources of rater bias. *Journal of Counseling Psychology* 1988;35:346–350.
- Hoffart A, Sexton H, Nordahl HM, Stiles TC. Connection between patient and therapist and therapist's competence in schema-focused therapy of personality problems. *Psychotherapy Research* 2005;15:409–419.
- Hoyt WT. Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods* 2000;5:64–86. [PubMed: 10937323]
- Hoyt WT, Kerns MD. Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods* 1999;4:403–424.
- Hoyt WT, Melby JN. Dependability of measurement in counseling psychology: An introduction to generalizability theory. *The Counseling Psychologist* 1999;27:325–352.
- Hu, LT.; Bentler, PM. Evaluating model fit. In: Hoyle, RH., editor. *Structural equation modelling: Concepts, issues and applications*. Sage; Thousand Oaks, CA: 1995. p. 77-99.
- Kline, RB. Principles and practice of structural equation modeling. Guilford Press; New York: 1998.
- Mahalik J, Hill CE, O'Grady KE, Thompson B. Rater bias in the Checklist of Psychotherapy Transactions-Revised. *Psychotherapy Research* 1993;3:47–56.
- Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin* 1988;103:391–410.
- Martino S, Ball SA, Nich C, Frankforter TF, Carroll KM. Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence* 2008;96:37–48. [PubMed: 18328638]
- McCarthy KS, Barber JP. The Multitheoretical List of Therapeutic Interventions (MULTI): Initial report. *Psychotherapy Research*. in press
- McHugo GJ, Drake RE, Teague GB, Xie H. Fidelity to assertive community treatment and client outcomes in the New Hampshire dual disorders study. *Psychiatric Services* 1999;50:818–824. [PubMed: 10375153]
- Miller WR. Bring addiction treatment out of the closet. *Addiction* 2007;102:863–869.
- Miller WR, Benefield B, Tonigan JS. Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology* 1993;61:455–461. [PubMed: 8326047]

- Miller WR, Mount KA. A small study of training in motivational interviewing: Does one workshop change therapist and client behavior? *Behavioural and Cognitive. Psychotherapy* 2001;29:457–471.
- Miller, WR.; Rollnick, S. *Motivational Interviewing: Preparing people for change*. Vol. 2nd ed. Guilford Press; New York: 2002.
- Miller WR, Sorensen JL, Selzer JA, Bringham GS. Disseminating evidence-based practices in substance abuse treatment: A review with suggestions. *Journal of Substance Abuse Treatment* 2006;31:25–39. [PubMed: 16814008]
- Miller WR, Yahne CE, Moyers TE, Martinez J, Pirritano M. A randomized trial of methods to help therapists learn motivational interviewing. *Journal of Consulting and Clinical Psychology* 2004;72:1050–1062. [PubMed: 15612851]
- Miller, WR.; Zweben, A.; DiClemente, CC.; Rychtarik, RG. *Motivational Enhancement Therapy manual: A clinical research guide for therapists treating individuals with alcohol abuse and dependence*. Vol. 2. National Institute on Alcohol Abuse and Alcoholism; Rockville, MD: 1992. Project MATCH Monograph Series
- Miller WR, Zweben J, Johnsen WR. Evidence-based treatment: Why, what, where, when, and how? *Journal of Substance Abuse Treatment* 2005;29:267–276. [PubMed: 16311179]
- Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods* 2006;38:542–547. [PubMed: 17186766]
- O'Malley SS, Foley SH, Rounsaville BJ, Watkins JT, Sotsky SM, Imber SD, et al. Therapist competence and patient outcomes in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology* 1988;56:496–501. [PubMed: 3198805]
- Perepletchikova F, Kazdin AE. Therapist integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice* 2005;12:365–383.
- Santa Ana EJ, Martino S, Ball SA, Nich C, Frankforter TL, Carroll KM. What is Usual about 'Treatment-as-Usual'? Data from Two Multisite Effectiveness Trials. *Journal of Substance Abuse Treatment*. in press
- Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. *American Psychologist* 1989;44:922–932.
- Shaw BF, Elkin I, Yamaguchi J, Olmstead M, Vallis TM. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology* 1999;67:837–846. [PubMed: 10596506]
- Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, Carroll KM. We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive behavioral therapy. *Journal of Consulting and Clinical Psychology* 2005;73:106–115. [PubMed: 15709837]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420–429. [PubMed: 18839484]
- Silove D, Parker G, Manicavasagar V. Perceptions of general and specific therapist behaviors. *Journal of Nervous and Mental Disease* 1990;178:292–299. [PubMed: 2338536]
- Waltz J, Addis ME, Koerner K, Jacobson N, S. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology* 1993;61:620–30. [PubMed: 8370857]
- Weiss DD, Marmar CR, Horowitz MJ. Do the ways in which psychotherapy process ratings are made make a difference? The effects of mode of presentation, segment, and rating format on interrater reliability. *Psychotherapy* 1988;25:44–50.
- Yadama GN, Pandey S. Effect of sample size on goodness-of-fit indices in structural equation models. *Journal of Social Services Research* 1995;20:49–70.

Table 1

Independent Tape Rater Scale item definitions for the fundamental and advanced MET strategies factors

Fundamental MET Strategies	Definition
Open Questions	Questions that elicit more than yes/no responses and encourage the client to talk about his/her problems, motivation, change efforts, and plans
Reflective Statements	Repeating (exact words), rephrasing (slight rewording), paraphrasing (e.g., amplifying the thought or feeling, use of analogy, making inferences) or making reflective summary statements about what the client said
Affirmations	Verbal reinforcement of the client's strengths, abilities, or efforts to change his/her behavior
Fostering a Collaborative Relationship	Statements or questions that convey that the therapy is a collaborative relationship in contrast to one where the therapist is in charge
Motivational Interviewing Style	Demonstration in the session of the therapist's supportive, warm, and non-judgmental stance, empathic sensitivity to the client's experiences, and attentiveness to the client's motivation for change
Advanced MET Strategies	
Client-centered Problem Discussion and Feedback	Explicit facilitation of a discussion about the problems for which the client entered treatment, as well as the provision of personalized feedback about the client's substance abuse and the evidence or indications of related problems in other life areas
Pros, Cons, and Ambivalence	Exploration of the positive and negative effects of the client's substance use, what might be gained and lost by abstinence or reduction in substance use, and reflections capturing the client's ambivalence about the change process
Heightening Discrepancies	Efforts made to heighten the internal conflict a client's continuing substance use poses to the client's self perceptions, life goals or values
Motivation for Change	Efforts made to elicit client discussion of change (self-motivational statements) through questions or comments designed to promote greater awareness/concern for the problem, increase intent/optimism to change, or encourage elaboration on a topic related to change
Change Planning	Discussion with a client about his or her readiness to prepare a change plan and facilitation of a discussion about the client's self-identified goals, steps for achieving those goals, supportive people available to help the client, what obstacles to the change plan might exist, and how to address impediments to change

Table 2

Percentage of sessions with MET strategies rated as present by therapists, supervisors, and observers

Items	Therapist	Supervisor	Observer	% Absolute Agreement
<u>Fundamental Strategies</u>				
1. Open-ended Questions	100	100	97	97
2. Reflective Statements	100	100	100	97
3. Affirmations of Strengths/Self-Efficacy	100	95	77	74
4. Fostering a Collaborative Relationship	99	85	49	37
5. Motivational Interviewing Style	100	100	91	92
<u>Advanced Strategies</u>				
1. Client-Centered Problem Discussion and Feedback	99	77	82	63
2. Pros, Cons, and Ambivalence	98	77	60	51
3. Heightening Discrepancies	97	79	24	25
4. Motivation for Change	99	91	65	60
5. Change Planning	93	67	50	38

Note. Absolute agreement is the percentage of the set of three ratings in which the raters agree that the item was either present or absent.

Table 3

Variance component estimates for MET adherence and competence

	Adherence	Competence
Rater Type (R)	.39	.47
Item Type within R	.14	.13
Tape (T)	.09	.12
RT	.00	.00
Residual	.38	.28
Generalizability Coefficient	.60	.63

Note: Raters were therapists, supervisors and observers for adherence and supervisors and observers for competence. Item types were fundamental and advanced MET strategies. Tapes were the individual recorded client sessions. The generalizability coefficient is relative. Competence scores reflect ratings of supervisors and observers only.

Table 4

Mean fundamental and advanced MET strategy adherence and competence ratings among therapists, supervisors, and observers

	Mean (SD)		n	F-test Statistic	p value	Significant Post Hoc Comparisons	
	Therapist (T)	Supervisor (S)				T vs. O	S vs. O
Adherence ^a							
Fundamental Strategies	5.2 (0.8)	5.0 (0.9)	351	30.3	.00	T > J	S > J
Advanced Strategies	4.6 (0.9)	3.6 (1.0)	351	127.4	.00	T > J	S > J
Competence ^b							
Fundamental Strategies	--	5.4 (1.0)	234	14.6	.00		
Advanced Strategies	--	4.9 (1.0)	233	1.1	.30		

Note. F-test statistic based on one-way ANOVA using 351 rated tapes (117 tapes × 3 raters), though sample sizes varied for competence ratings that are made only when an item occurs within a session; to reduce Type I error p-values ≥ .025 (.05/2) are significant for tests involving fundamental and advanced MET strategies; Tukey's honestly significant difference test was used for post hoc analyses for adherence ratings comparing group means.

^a Adherence rating means and standard deviations are based on a 7-point Likert adherence scale: 1=not at all, 2=a little (once), 3=infrequently (twice), 4=somewhat (3-4 times), 5=quite a bit (5-6 times), 6=considerably (> 6 times/more depth in interventions), 7=extensively (high frequency/characterizes entire session).

^b Competence rating means and standard deviations are based on a 7-point Likert competence scale: 1= very poor, 2=poor, 3=acceptable, 4=adequate, 5=good, 6=very good, 7 = excellent.