Published in final edited form as:

Psychol Assess. 2013 March ; 25(1): 306–312. doi:10.1037/a0030815.

Correspondence of Motivational Interviewing Adherence and Competence Ratings in Real and Role-Played Client Sessions

Suzanne E. Decker

New England Mental Illness Research Education and Clinical Centers / VA Connecticut Healthcare System, Department of Psychiatry, Yale University School of Medicine

Kathleen M. Carroll, Charla Nich, and Monica Canning-Ball

Department of Psychiatry, Yale University School of Medicine

Steve Martino

Department of Psychiatry, Yale University School of Medicine, VA Connecticut Healthcare System

Abstract

Treatment integrity ratings (adherence and competence) are frequently used as outcome measures in clinician training studies, drawn from recorded real client or role-played client sessions. However, it is unknown whether clinician adherence and competence are similar in real client and role-played sessions or whether real and role-play clients provide similar opportunities for skill demonstration. This study examined the correspondence of treatment adherence and competence ratings obtained in real client and role-played sessions for 91 clinicians trained in Motivational Interviewing (MI), using data from a multi-site trial examining three methods of clinician training (Martino et al., 2011). Results indicated overall poor integrity rating correspondence across the two session types, as indicated by weak correlations ($r = .05-.27$). Clinicians were rated significantly more MI adherent overall and specifically used more advanced MI strategies in role-played than real client sessions at several assessment time points ($d = 0.36, 0.42$). Real clients, in comparison to the role-play actor, demonstrated greater motivation at the beginning of the session ($d = 1.09$), discussion of unrelated topics ($d = 0.70$), and alliance with the clinician ($d = 0.72$). These findings suggest that MI integrity rating data obtained from real client and role-played sessions may not be interchangeable. More research is needed to improve the procedures and psychometric strength of treatment integrity assessment based on role-played sessions.

Keywords

dissemination; psychotherapist training; adherence; competence; motivational interviewing

One of the greatest challenges of disseminating empirically supported treatments (EST) is training clinicians to implement them with integrity, namely, with adequate levels of adherence and competence (McHugh & Barlow, 2010). Independent rating of audio or video recorded sessions using observer rating scales (e.g., Carroll et al., 2000) has become the gold standard for assessing treatment integrity in clinical and clinician training trials (Herschell, Kolko, Baumann, & Davis, 2010). Typically, independent raters assess treatment integrity in recorded sessions with real clients drawn from the clinicians' caseloads (e.g., Forsberg, Forsberg, Lindqvist, & Helgason, 2010) or role-played sessions using actors who follow

standardized scripts (e.g., Baer et al., 2004; Sholomskas et al., 2005). Some studies have combined both assessment methods (e.g., Miller et al., 2004). However, the extent to which these different methods are comparable is unclear. This study examines the correspondence of clinician motivational interviewing (MI; Miller & Rollnick, 2002) adherence and competence ratings obtained using real and role-played client sessions during a clinician training trial (Martino et al., 2011).

Benefits and drawbacks exist for each treatment integrity assessment method. Real client sessions allow for assessment of clinicians' implementation of EST in their practice settings (Herschell et al., 2010) and how implementation may vary with clients who present differently (Imel, Baer, Martino, Ball, & Carroll, 2011). Drawbacks include: 1) fluctuation in clinicians' performance due to case difficulty that may not accurately show the trajectory for how well clinicians learn EST (Imel et al., 2011); 2) clinicians' reluctance to obtain recordings because it is a non-routine practice that requires express client consent (Baer et al., 2004); and 3) work site interruptions (e.g., clinical crises, staffing shortages) that hamper efforts to gather recorded real client sessions (Forsberg et al., 2010).

Advantages of standardized role-play assessment are: 1) less variability in clinician performance associated with different client presentations (Baer et al., 2004); 2) sessions can be recorded immediately after training, reducing the potential for lost data (e.g., Miller et al., 2004); and 3) development of specific scenarios to allow clinicians to demonstrate key skills (Sholomskas et al., 2005) or to represent a standard clinical encounter (Baer et al., 2004). However, role-play actors may not respond to clinicians in the manner real clients might act in that they may stick to a script (Miller et al., 2004) or not represent the types of clients clinicians are likely to treat (e.g., level of motivation). Thus, it is unclear whether clinician performance in role-play sessions provides an accurate estimate of whether clinicians can perform an EST adequately in clinical settings.

No previous research has examined the correspondence of treatment integrity ratings obtained from real and role-played sessions gathered at similar points in time. In a recent study (Martino et al., 2011), we examined the effectiveness of expert-led and train-the-trainer strategies, compared to a self-study approach, for training clinicians in MI. Primary outcomes were MI adherence and competence in recorded real client and standardized role-play sessions. In the present report, we compared the clinicians' MI adherence and competence in recorded real and role-played client sessions simultaneously gathered at each of four assessment points (baseline, after workshop training, after supervision, and at 12-week follow-up) in order to determine the correspondence of these two different treatment integrity assessment methods.

Method

Participants

Ninety-two clinicians from 12 outpatient substance abuse programs in the State of Connecticut participated in the original study (see Martino et al., 2011 for demographic details). All participants were required to be employed at least 20 hours per week, treat English-speaking substance-using clients, and provide written informed consent as approved by the Yale University School of Medicine Human Investigation Committee. Of the 92 participants, 91 (99%) provided both a real and role-played client session for at least one of the four assessment time point and provided the data for the present report. Retention rates were good, though collection of sessions varied (72–98%) depending on clinician compliance, operator error, and equipment failure. Sample sizes per time points were: baseline = 84; post-workshop = 75; post-supervision = 76; and 12-week follow-up = 65, with

26 observations missing real client sessions and 18 observations missing role-play sessions, for a total of 300 real and role-played session pairs across time points.

Procedures

Clinician Training—Details about the original study's aims, methods, and results have been published previously (Martino et al., 2011). Treatment programs were randomly assigned to one of three training conditions. Clinicians received the training strategy to which their program had been randomly assigned (self-study = 31; expert-led = 32; train-the-trainer = 29).

Real Client Sessions—Clinicians audiotaped 40-minute sessions in which they conducted MI with clients who had substance use problems. Clinicians, not research staff, selected clients based on clinical judgment of the suitability of substituting MI for their standard practice and the client's willingness to be audiotaped (signed consent obtained). The research assistant gave the clinician written instructions that asked the clinician to motivate the selected client to change his or her substance use. Audiotaping was used to keep the recording procedures simple for the clinicians. Client sessions were obtained within 2 weeks of each assessment point.

Role-Played Client Sessions—A role-play scenario of an initial client session using a female client-actor was developed to control for program-specific populations or variability in clinical presentations¹. A composite client scenario was created to be credible across programs and reviewed by clinical directors before a final version was written (i.e., a 30-year-old woman on probation for a drunken driving conviction with a history of alcohol and cannabis abuse; she was ambivalent about changing her substance use). A brief clinical intake form was developed that summarized the client's presenting problems, demographic information, and preliminary diagnoses based on the presumption that the client had completed the program's intake and was meeting with the assigned clinician for the first time. Immediately before the session, the research assistant handed the clinician instructions that informed the clinician he or she had been assigned to treat this client and asked the clinician to motivate the client to change her substance use. The actor was instructed to let her experience of the clinician guide her responses. If she felt listened to, understood, and supported by the clinician, the actor was told to be more open and collaborative. If these conditions were not met, the actor was told to become less engaged with the clinician and more resistant in style. All sessions were 40 minutes in length, videotaped, and obtained within two weeks of each assessment point. Videotaping was used to record nonverbal clinician behavior that might inform MI adherence and competence ratings, since these behaviors could not be assessed from the audiotaped real client sessions.

Rater Training—Twelve raters rated the real and role-played sessions using the Independent Tape Rater Scale (ITRS). Raters received approximately 44 hours of ITRS training in seminars and through practicing rating tapes with expert feedback (see Martino et al., 2011 for details).

Measures

Independent Tape Rater Scale (ITRS)—The ITRS (Martino, Ball, Nich, Frankforter, & Carroll, 2008) assesses community program clinicians' adherence and competence using MI and other strategies inconsistent with MI (e.g., direct confrontation) or common to drug

¹More information on the development of the role-play scenario, actor training, and actor presentation is available in the extended report version.

counseling (e.g., assessing substance use). For each strategy, raters use a 7-point Likert-type scale reflecting the frequency or extent (adherence; 1 = not at all, to 7 = extensively) and the skill with which it is used (competence; 1 = very poor, to 7 = excellent). Prior confirmatory factor analyses of the ITRS items supported a two-factor solution for MI consistent items reflecting five fundamental MI strategies (e.g., open questions; affirmations) and five advanced MI strategies (e.g., developing discrepancies between substance use and personal values, goals, and self-perceptions) (Martino et al., 2008). Factor scores are calculated using the means of the related five MI strategy items to yield fundamental adherence, fundamental competence, advanced adherence, and advanced competence scores². A clinician who conducts a session in which he or she achieves at least half of the 10 MI-consistent items rated average or above in terms of adherence and competence is deemed adequately performing MI, the same criterion used in clinical effectiveness trials of MI (Ball et al., 2007; Carroll et al., 2006). ITRS item test scores have shown good to excellent inter-rater reliability (Martino et al., 2011; Martino et al., 2008). Intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979) were calculated from ratings provided by all raters from an identical set of 18 calibration tapes (9 real and 9 role-played client sessions) randomly selected from the larger pool of session tapes, with ICC estimates for the MI factors as follows: fundamental MI adherence = .88, advanced MI adherence = .87; fundamental MI competence = .87, advanced MI competence = .68). ICCs calibrated from the separate real and role-played client session samples were similar. ITRS test scores for items on client in-session behavior demonstrated excellent inter-rater reliability: motivation at beginning of session (ICC = .91), discussion of topics unrelated to treatment (ICC = .90), difficulty understanding the clinician (ICC = .85), and alliance with clinician (ICC = .93).

Data Analysis

We estimated the degree of correspondence between the clinicians' demonstration of MI integrity in real and role-played client sessions in several ways. We calculated Pearson correlations for the integrity ratings of real and role-played sessions for each of the four respective MI strategy factor scores derived from the ITRS (i.e., fundamental and advanced MI adherence and competence). Next, we used mixed-effects regression models to examine the effect of session type (real, role-played) on MI integrity ratings relative to other fixed factors (training condition [self-study, expert-led, train-the-trainer], strategy [fundamental, advanced], dimension [adherence, competence]) and their interactions, with clinician as a random factor. Significant effects for session type or for its interaction with other factors would suggest poor overall MI integrity rating correspondence between real and role-played sessions. Post-hoc comparisons were conducted on significant effects. Effect sizes for post-hoc comparison tests were estimated using Raudenbush and Liu's formula (2001). Because the main outcomes paper indicated that clinician performance in some training conditions improved over time (Martino et al., 2011), we present mixed-effects regression models run separately at each assessment time point as well as a model with all data points combined over time.

We assessed agreement between real and role-played sessions in meeting the criterion of adequate MI performance by calculating kappas, absolute agreement, and the observed proportion of positive and negative agreement (clinician met standard; clinician did not meet standard). Chi-square analyses were used to examine differences between session type in the proportion of clinicians who met this criterion. Finally, we examined client behaviors to understand potential differences in the clinical presentation of the real clients and role-play actor. We used a doubly multivariate analysis of variance test (Tabachnick & Fidell, 2007)

²More information on the ITRS, including a list of items, may be found in Martino et al. (2008). The measure is available on request from the senior author.

with client behavior variables as the multivariate dependent variable, session type and training condition as between-subjects independent variables, and assessment time point as a within-subjects independent variable.

Results

Correspondence of MI Integrity

Table 1 presents the means and standard deviations for MI strategy integrity ratings for real and role-played client sessions combined across training conditions. Mean MI adherence scores indicated “a little” (2) to “somewhat” (4) frequent use of MI strategies, while mean MI competence scores ranged from “adequate” (4) to “good” (5).³ Weak positive Pearson correlations between real and role-played session scores for fundamental MI adherence and competence were obtained at several assessment time points and for all data combined across time. No correlations reached significance for the advanced skills.

Table 2 shows the results of the mixed-effects regression models that examined the effect of session type on MI integrity ratings in relationship to other fixed (training condition, strategy, dimension) and random (clinician) factors for each assessment time point. As reported in the main trial (Martino et al., 2011), significant main effects occurred for condition post-workshop, post-supervision, and 12-week follow-up. At all time points, significant main effects were present for strategy (fundamental strategies were used on average more often than advanced strategies), and dimension (mean competence ratings were greater than adherence ratings); these factors significantly interacted with each other (e.g., at baseline, fundamental adherence was greater than advanced adherence, but no such differences in competence were found).

Significant main effects for session type occurred at the post-workshop assessment point, with role-played client mean integrity ratings greater than those obtained from real clients. Session type interacted with strategy, with clinicians using significantly more advanced strategies in role-played than real client sessions, at post-workshop (estimated $M = 3.76$ vs. 3.35 , $SE = 0.08$, $d = 0.36$) and 12-week follow-up (estimated $M = 3.67$ vs. 3.46 , $SE = 0.08$, $d = 0.42$). In addition, session type interacted with dimension such that mean adherence ratings were greater in role-played than real client sessions at baseline (estimated $M = 3.36$ vs. 3.19 , $SE = 0.07$, $d = -0.32$), post supervision (estimated $M = 3.71$ vs. 3.56 , $SE = 0.08$, $d = -0.46$), and 12-week follow-up (estimated $M = 3.62$ vs. 3.51 , $SE = 0.08$, $d = -0.12$).

Correspondence of meeting criterion of adequate MI performance—Table 3 shows several rating concordance indices for real client and role-played sessions. Statistically significant but low magnitude kappas (Fleiss, 1981) were obtained at baseline, post-supervision, and for the entire sample, suggesting poor correspondence between real and role-played sessions in meeting the criterion level of performance. Similarly, the percentage of agreement between the type of sessions meeting the performance standard was low to moderate. Average proportional agreement about meeting (p_{pos}) or not meeting (p_{neg}) the standard, analogous to sensitivity and specificity (Cicchetti & Feinstein, 1990), indicated the low kappa values were not disproportionately due to agreement on meeting the standard or not. The chi-square for differences in meeting the adequate standard was significant for data combined across all assessment time points, with clinicians more likely to meet the standard in role-played sessions.

³An extended version of this Brief Report, including mean integrity ratings and correlations among them separated by training condition, is available upon request from the authors.

Client Behavior

We examined in-session client behaviors to determine whether variation in client behavior across session types might contribute to the poor correspondence. The MANOVA showed that as a whole client behavior differed significantly by session type ($F(4, 93) = 42.90$, Wilks Lambda = .35, $p = .000$); no effects were detected for training condition, time, or interactions. Univariate F tests indicated that, in comparison to the role-play actor's client portrayal, real clients were significantly more motivated at the session beginning (estimated means for real clients in self-study, expert-led, and train-the-trainer conditions, respectively, were 8.26, 8.76, and 8.92; for role-play clients, 6.68, 6.47, and 6.23; $F(1, 96) = 103.15$, $p = .000$), likely to discuss unrelated topics (estimated means for real clients in self-study, expert-led, and train-the-trainer conditions, respectively, were 4.29, 3.79, and 3.19; for role-play clients, 2.26, 2.37, and 2.23, $F(1, 96) = 46.14$, $p = .000$), and allied with their clinicians (estimated means for real clients in self-study, expert-led, and train-the-trainer conditions, respectively, were 8.82, 9.26, and 9.00; for role-play clients, 7.66, 7.89, and 7.23, $F(1, 96) = 41.40$, $p = .000$). Effect sizes were large for motivation ($d = 1.09$) and medium for alliance and unrelated topics ($d = 0.70, 0.72$). No session type differences occurred for clients' capacity to understand the clinician.

Discussion

Data from this study calls into question the interchangeability of treatment integrity ratings derived from real client versus role-played sessions, in that clinicians across training conditions and assessment time points did not demonstrate the same ability to use MI in contiguous real client and role-played sessions. With few exceptions (e.g., weak positive correlations for fundamental MI skills), this was found among several measures that examined the rating correspondence of fundamental and advanced MI adherence and competence ratings and the degree to which clinicians met a criterion level of adequate MI performance.

A possible reason for the low correspondence among MI integrity ratings is that the clients' presentations within the two types of sessions differed. Real clients were independently rated as more motivated to change their substance use and more allied with the clinicians than the role-play actor. These types of client behaviors have been shown to influence clinician MI integrity: Imel et al. (2011) found that increased client motivation was associated with less clinician use of MI strategies, while Crits-Christoph et al. (2009) demonstrated that higher levels of client-rated alliance were associated with increased MI adherence. While the direction of influence is not clear (e.g., use of MI strategies may foster alliance), clients who are more engaged with the clinician may facilitate more effective demonstration of MI skills.

A second reason for the poor rating correspondence is that role-play assessments often are designed to target specific clinician abilities. If these targeted abilities differ from those needed to enhance client motivation in actual sessions, then some integrity rating discordance would be expected. In our study, we purposely trained the actor to present herself as ambivalent about changing her substance use. We anticipated that clinicians might be prone to invite more motivated or engaged real clients to participate in this training trial to increase the chance they could demonstrate their best MI skills (Miller et al., 2004). As described above, this selection bias likely occurred, and the clinicians appeared to adjust their MI practice to accommodate the need to use more MI consistent strategies, particularly advanced ones, in the role-played sessions to elicit motivations for change from an actor who was more ambivalent than the actual clients.

A third possible explanation for poor rating correspondence may be the difference in novelty between the real and role-played sessions. In an effort to standardize the test of MI integrity, we repeated the same role-play assessment without varying the clinical circumstances. However, this procedure risks practice effects that may inflate MI integrity ratings in the role-played sessions (Helitzer et al., 2011). Moreover, as found in our study, repeated use of the same role-played session may diminish the formation of therapeutic alliance in that actors may be less responsive to clinician behavior if they stick to a script (Miller et al., 2004); clinicians similarly may succumb to scripted behavior in repeated role-plays with the same actor and clinical scenario. Future clinician training studies that use role-play assessment should develop alternate forms in which clinicians would be presented with different cases that target the demonstration of similar therapeutic skills. For example, Baer and colleagues (2012) have showed equivalently good rater reliability of test scores, as well as construct and criterion-related validity of test score interpretation, on alternative versions of a computer-delivered simulated patient interview to assess MI integrity.

Given the overall poor correspondence between real client and role-played sessions, which type of integrity assessment method should be used to evaluate clinicians' ability to implement EST? One suggestion would be to only use real client sessions until the procedures and psychometric properties for role-played sessions are better established. However, more research is still needed to validate the use of role-played client session integrity assessment in real world settings and to examine its relationship to client outcomes. In this regard, the field has no empirically established integrity rating performance benchmarks derived from any assessment method for any EST.

Nonetheless, if the main objective is to determine the degree to which clinicians are adherent and competent in their performance of key interventions of an EST, then role-played sessions might be useful. Our role-played session was designed to maximize opportunities for applying advanced MI strategies, and we found clinicians used these strategies more frequently in the role-played sessions. Other investigators have taken the same approach. For example, Sholomskas et al. (2005) used brief structured role-plays with client-actors to assess the ability of clinicians to conduct several specific and essential components of cognitive behavioral therapy. In such multi-session treatments, capturing instances in which clinicians should apply key interventions might be more difficult without targeted role-played exercises. Perhaps for this reason, role-played session assessment occurs frequently in training trials involving cognitive behavioral therapy (Herschell et al., 2010).

This study's limitations include (1) the use of only one assessment of real and role-play MI integrity at each time point; (2) clinicians selecting the real client sessions, which may have resulted in selection of particularly engaged or allied clients and does not permit us to determine which session type produces a more valid assessment of clinician skill; (3) the conflation of recording method (audio vs. video) with session type; (4) the use of only one role-play actor and clinical scenario; (5) attrition over time for recorded real and role-play client sessions; and (6) lack of data about relationship of MI integrity to client outcomes for both session types.

This study directly compared treatment integrity ratings obtained from two common sources, real client and role-played sessions. Poor rating correspondence was demonstrated between the two assessment approaches for fundamental and advanced MI adherence and competence dimensions and in the percentage of clinicians who met a clinical trials criterion level of adequate MI performance within sessions. These results suggest that the two assessment formats should not be used interchangeably to assess changes in clinician MI integrity over time, as has been done in prior training studies (e.g., Miller et al., 2004). More research is needed to develop the procedures and psychometric strength of integrity

assessment based on role-played sessions, including empirically linking benchmarks of performance to client outcomes.

Acknowledgments

Writing of this manuscript was supported by the Office of Academic Affiliations, Advanced Fellowship in Mental Illness Research and Treatment, Department of Veterans Affairs, the Veterans Affairs Connecticut Healthcare System, and the Department of Veterans Affairs New England Mental Illness Research, Education, and Clinical Center (MIRECC). This study was funded by the US National Institute on Drug Abuse (R01 DA16970 awarded to Steve Martino, with additional support provided by R01 DA023230, U10-DA013038 and P50-DA09241).

References

- Baer JS, Carpenter KM, Beadnell B, Stoner SA, Ingalsbe MH, Hartzler B, Drager Z. Computer assessment of simulated patient interviews (CASPI): Psychometric properties of a web-based system for the assessment of motivational interviewing skills. *Journal of Studies on Alcohol and Drugs*. 2012; 73:154–164. [PubMed: 22152673]
- Baer JS, Rosengren DB, Dunn CW, Wells EA, Ogle RE, Hartzler B. An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*. 2004; 73:99–106. [PubMed: 14687964]
- Ball SA, Martino S, Nich C, Frankforter TL, Van Horn D, Crits-Christoph P, Carroll KM. Site matters: Multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Counseling and Clinical Psychology*. 2007; 75:556–567.
- Carroll KM, Ball SA, Nich C, Martino S, Frankforter TL, Farentinos C, Woody GE. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: A multisite effectiveness study. *Drug and Alcohol Dependence*. 2006; 81:301–312. [PubMed: 16169159]
- Carroll KM, Nich C, Sifry R, Frankforter T, Nuro KF, Ball SA, 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]
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*. 1990; 43:551–558. [PubMed: 2189948]
- Crits-Christoph P, Gallop R, Temes CM, Woody G, Ball SA, Martino S, Carroll KM. The alliance in motivational enhancement therapy and counseling as usual for substance use problems. *Journal of Counseling and Clinical Psychology*. 2009; 77:1125–1135.
- Fleiss, JL. *Statistical methods for rates and proportions*. 2nd ed.. John Wiley; New York: 1981.
- Forsberg L, Forsberg LG, Lindqvist H, Helgason AR. Clinician acquisition and retention of Motivational Interviewing skills: A two-and-a-half-year exploratory study. *Substance Abuse Treatment, Prevention, and Policy*. 2010; 5
- Helitzer DL, Lanoue M, Wilson B, de Hernandez BU, Warner T, Roter D. A randomized controlled trial of communication training with primary care providers to improve patient-centeredness and health risk communication. *Patient Education and Counseling*. 2011; 8:21–29. [PubMed: 20219315]
- Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendation. *Clinical Psychology Review*. 2010; 30:448–466. [PubMed: 20304542]
- Imel ZE, Baer JS, Martino S, Ball SA, Carroll KM. Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence*. 2011; 115:229–236. [PubMed: 21227601]
- Martino S, Ball SA, Nich C, Canning-Ball M, Rounsaville BJ, Carroll KM. Teaching community program clinicians motivational interviewing using expert and train-the-trainer strategies. *Addiction*. 2011; 106:428–441. [PubMed: 20925684]
- Martino S, Ball SA, Nich C, Frankforter TL, Carroll KM. Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence*. 2008; 96:37–48. [PubMed: 18328638]

- McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*. 2010; 65:73–84. [PubMed: 20141263]
- Miller, WR.; Rollnick, S. *Motivational interviewing: Preparing people to change addictive behavior*. 2nd ed.. Guilford Press; New York: 2002.
- Miller W, Yahne C, Moyers T, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*. 2004; 72:1050–1062. [PubMed: 15612851]
- Raudenbush SW, Liu X-F. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*. 2001; 6:387–401. [PubMed: 11778679]
- Sholomskas DE, Syracuse-Siewart 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; 2:420–428. [PubMed: 18839484]
- Tabachnick, BG.; Fidell, LS. *Using multivariate statistics*. 5th ed.. Pearson Education; Boston: 2007.

Table 1

Mean MI integrity ratings for real and role-played sessions and correlations among them

Strategy	Assessment Time Point	N	Real M (SD)	Role M (SD)	r	p
Fundamental Adherence	Baseline	84	3.93 (0.94)	3.94 (0.84)	.32**	.00
	Post-workshop	75	4.23 (1.00)	4.39 (0.80)	.18	.12
	Post-supervision	76	4.33 (1.02)	4.33 (0.70)	.24*	.04
	12-week Follow-up	65	4.31 (0.89)	4.29 (0.78)	.25*	.04
	All	300	4.19 (0.98)	4.23 (0.80)	.27**	.00
Fundamental Competence	Baseline	84	3.97 (0.75)	3.91 (0.90)	-.01	.91
	Post-workshop	75	4.29 (0.97)	4.44 (0.85)	.26*	.03
	Post-supervision	76	4.58 (0.84)	4.49 (0.89)	.22	.05
	12-week Follow-up	65	4.48 (1.00)	4.40 (0.83)	.22	.07
	All	300	4.31 (0.91)	4.30 (0.90)	.23**	.00
Advanced Adherence	Baseline	84	2.45 (0.86)	2.79 (1.02)	-.01	.95
	Post-workshop	75	2.58 (1.05)	3.17 (0.88)	-.03	.82
	Post-supervision	76	2.80 (0.97)	3.08 (0.96)	.15	.19
	12-week Follow-up	65	2.53 (0.87)	3.00 (0.95)	.15	.23
	All	300	2.59 (0.95)	3.00 (0.96)	.07	.20
Advanced Competence	Baseline	80	4.05 (0.80)	3.97 (0.91)	-.13	.24
	Post-workshop	73	4.12 (0.84)	4.34 (0.74)	.12	.31
	Post-supervision	76	4.47 (0.71)	4.34 (0.74)	.10	.40
	12-week Follow-up	62	4.45 (0.78)	4.38 (0.77)	.05	.69
	All	291	4.26 (0.80)	4.25 (0.81)	.05	.38

Note. Significant values of

*
 $p < .05$,**
 $p < .01$ are bolded.

Table 2

Mixed-effects regression models on MI integrity ratings

Source	N. df	Baseline			Post-workshop			Post-supervision			12-week Follow-up		
		Den. df	F	p	Den. df	F	p	Den. df	F	p	Den. df	F	p
Intercept	1	80.91	5285.31	.00	71.90	5057.12	.00	73.00	4474.06	.00	61.71	4332.30	.00
Condition	2	80.90	2.42	.10	71.91	8.05	.00	73.00	6.36	.00	61.71	8.26	.00
Session Type	1	568.32	0.56	.45	507.08	18.60	.00	516.00	0.00	.95	435.95	1.68	.20
Strategy	1	568.31	102.66	.00	507.08	156.27	.00	516.00	166.75	.00	435.95	138.67	.00
Dimension	1	568.31	126.19	.00	507.08	122.54	.00	516.00	201.61	.00	435.95	177.33	.00
Condition × Session Type	2	568.32	4.90	.01	507.08	0.70	.50	516.00	1.91	.15	435.95	3.23	.04
Condition × Strategy	2	568.30	0.33	.72	507.08	2.35	.10	516.00	0.22	.81	435.95	0.82	.44
Condition × Dimension	2	568.30	1.27	.28	507.08	1.41	.25	516.00	0.00	1.00	435.95	1.90	.15
Session Type × Strategy	1	568.32	1.73	.19	507.08	4.37	.04	516.00	1.49	.22	435.95	3.95	.05
Session Type × Dimension	1	568.32	3.81	.05	507.08	2.30	.13	516.00	6.18	.01	435.95	4.18	.04
Strategy × Dimension	1	568.29	126.92	.00	507.09	103.41	.00	516.00	121.22	.00	435.97	137.07	.00
Condition × Session Type × Strategy	2	568.32	0.25	.78	507.08	1.29	.28	516.00	1.13	.32	435.95	0.51	.60
Condition × Session Type × Dimension	2	568.32	1.34	.26	507.08	0.78	.46	516.00	2.01	.13	435.95	0.56	.57

Note. N. = numerator. Den. = denominator. Condition = self-study, expert-led training, train-the-trainer.

Session type = real vs. role-play session. Strategy = fundamental vs. advanced. Dimension = adherence vs. competence.

Table 3
Concordance between real and role-played sessions meeting criterion for adequate MI

Assessment Time Point	Percent meeting criterion			Percent agreement			95% CI	χ^2
	Real	Role	κ	Abs.	pos.	Neg.		
Baseline	31	38	.21*	64	17	48	.48 [.29, .67]	.72 [.60, .84]
Post-workshop	43	57	.14	56	28	28	.56 [.39, .73]	.56 [.41, .71]
Post-supervision	57	58	.38**	70	42	28	.74 [.61, .87]	.65 [.49, .81]
12-week Follow-up	45	55	-.00	49	25	25	.49 [.31, .67]	.49 [.31, .67]
All	43	52	.21***	60	28	33	.58 [.50, .66]	.62 [.55, .69]

Note.

* $p < .05$,

**

$p < .05$,

$p < .001$

^a Indicates observed proportion of positive agreement, calculated as two times agreement on meeting standard divided by sum of meeting standard based on client tapes and meeting standard based on role-play tape.

^b Indicates observed proportion of negative agreement, calculated as two times agreement on failure to meet standard divided by sum of failing to meet standard based on client tapes and failing to meet standard based on role-play tapes.