Published in final edited form as:

J Consult Clin Psychol. 2012 June ; 80(3): 450–464. doi:10.1037/a0028176.

Training Substance Abuse Clinicians in Motivational Interviewing Using Live Supervision via Tele-conferencing

Jennifer L. Smith

Department of Psychiatry, Columbia University

Kenneth M. Carpenter

Department of Psychiatry, Columbia University

Paul C. Amrhein

Psychology Department, Montclair State University

Adam C. Brooks

Department of Psychiatry, Columbia University

Deborah Levin

Department of Psychiatry, Columbia University

Elizabeth A. Schreiber

Department of Psychiatry, Columbia University

Laura A. Travaglini

New York State Psychiatric Institute

Mei-Chen Hu

Department of Psychiatry, Columbia University

Edward V. Nunes

Department of Psychiatry, Columbia University

Abstract

Background—Training through traditional workshops is relatively ineffective for changing counseling practices. Tele-conferencing Supervision (TCS) was developed to provide remote, live supervision for training motivational interviewing (MI).

Method—97 community drug treatment counselors completed a 2-day MI workshop and were randomized to: live supervision via tele-conferencing (TCS; $n=32$), standard tape-based supervision (Tape; $n=32$), or workshop alone (Workshop; $n=33$). Supervision conditions received 5 weekly supervision sessions at their sites using actors as standard patients. Sessions with clients were rated for MI skill with the Motivational Interviewing Treatment Integrity (MITI) coding system pre-workshop and 1, 8, and 20 weeks post-workshop. Mixed effects linear models were used to test training condition on MI skill at 8 and 20 weeks.

Results—TCS scored better than Workshop on the MITI for Spirit (mean difference = 0.76; $p < .0001$; $d = 1.01$) and Empathy (mean difference = 0.68; $p < .001$; $d = 0.74$). Tape supervision fell between TCS and Workshop, with Tape superior to Workshop for Spirit (mean difference = 0.40; $p < .05$). TCS was superior to Workshop in reducing MI non-adherence and increasing MI

Correspondence concerning this article should be sent to Dr. Jennifer L. Smith, NYS Psychiatric Institute, 1051 Riverside Drive, Box 120, New York, NY 10032. smitjen@nyspi.columbia.edu.

Adam C. Brooks is now at The Treatment Research Institute in Philadelphia, Pennsylvania. Elizabeth Schreiber is now at Connection Counseling in Madison, Wisconsin.

adherence, and was superior to Workshop and Tape in increasing the reflection to question ratio. Tape was superior to TCS in increasing complex reflections. Percentage of counselors meeting proficiency differed significantly between training conditions for the most stringent threshold (Spirit and Empathy scores ≥ 6), and were modest, ranging from 13% to 67%, for TCS and Tape.

Conclusion—TCS shows promise for promoting new counseling behaviors following participation in workshop training. However, further work is needed to improve supervision methods in order to bring more clinicians to high levels of proficiency and facilitate the dissemination of evidence-based practices.

Keywords

motivational interviewing; substance abuse; clinician training

The dissemination of new, evidence-based treatment methods into widespread clinical practice is a substantial challenge for the medical and mental health fields generally, and addiction treatment in particular (Institute of Medicine, 1998). Continuing education programs have traditionally used printed materials and didactic conferences for disseminating treatment advances. However, evidence suggests that these traditional approaches to training are relatively ineffective at changing actual medical or counseling practices. Training procedures that include interactive role-plays relevant to the clinical context allow opportunities for further practice, and provide feedback and reminders are more likely to enhance clinical skills in medical settings (Grol, 2001; Grol & Grimshaw, 2003). Similarly, Motivational interviewing (MI) workshop training procedures that include practice and feedback have been shown to increase the MI skill levels of counselors (Madison, Loignon, & Lane, 2009; Walters, Matson, Baer, & Ziedonis, 2005). Unfortunately, skill levels have been demonstrated to diminish after workshop training, indicating workshop training alone does not produce sustained improvement in MI skillfulness (Baer et al., 2004; Miller & Mount, 2001; Walters et al., 2005). Further, following workshop training, counselors tend to rate their MI proficiency more positively than their objectively assessed skill ratings (Miller & Mount, 2001). These findings suggest that training workshops may successfully expose clinicians to new treatment approaches and increase their confidence in the use of these techniques, but may not promote long-term proficiency, an important outcome that has direct implications for improving clinical practice. Thus, alternative training methods are needed to facilitate a more robust transfer of new treatment strategies to the clinical context.

Accordingly, studies have begun to examine supervision offered after workshop training (Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; Moyers et al., 2008; Martino, Ball, et al., 2011), and generally support the conclusion that methods involving objective feedback and coaching will improve acquisition of MI skill (Miller & Rose, 2009). The importance of approximating a counselor's clinical context during MI training has been emphasized (Miller & Rollnick, 2002), and the evidence suggests that feedback should be based on ratings of actual therapy sessions (Miller et al., 2004; Miller & Rose, 2009). This has generally been achieved through audiotaping of clinical encounters, which provides a direct window into a clinician's counseling style, and facilitates remote supervision, an important feature for disseminating empirically based treatment strategies when time and distance limit the availability of expert trainers. However, audiotaping also has disadvantages including limited adherence and tape selection issues such as counselors may not make tapes, or may select sessions that went relatively well, limiting the ability to address problems with skill (Miller et al., 2004). Further, the use of audio recordings entails a delay between the clinical encounter and the provision of feedback to the counselor; the tape must be mailed and then reviewed by the supervisor before discussion and feedback can take place. This misses the

opportunity for immediate feedback, a potentially powerful intervention for helping individuals learn new behavioral skills (Balcazar, Hopkins, & Suarez, 1986).

Independent lines of evidence highlight three training parameters that may be particularly useful for guiding the learning of new counseling skills. First, immediate feedback is an important component in training programs that aim to bring behaviors to a particular performance level (Balcazar et al., 1986). Thus, supervisory practices that provide feedback to counselors closer in time to their clinical encounter may help promote the acquisition and retention of new counseling skills (Kivlighan, Angelone, & Swafford, 1991). Second, practice sessions distributed over time are more likely to produce lasting skill acquisition than training sessions of a longer duration but conducted over a shorter period of time (e.g., workshop training) (Donovan & Radosevich, 1999; Prescott et al., 2002). Third, differential reinforcement of specific behaviors from amongst a larger repertoire can be particularly effective at shaping new skills (Skinner, 1953). Differential reinforcement procedures may be particularly effective for developing proficiency in specific MI skills, while simultaneously reducing the use of counseling strategies that are incompatible with an MI style. Reducing the frequency of MI non-adherent (MINA) behaviors is of particular importance, since evidence suggests that therapist MINA behavior can increase client resistance and negatively influence treatment outcomes (Apodoca & Longabaugh, 2009; Guame, Bertholet, Faouzi, Gmel, & Daeppen, 2010; Guame, Gmel, Faouzi, & Daeppen, 2009; Miller, Benefield, & Tonigan, 1993; Moyers & Martin, 2006).

Tele-conferencing supervision (TCS) was designed to address the limitations of workshop training and post-workshop supervision based on recordings of clinical encounters, by incorporating the principles of immediate feedback, distributed practice, and differential reinforcement. TCS provides real-time, live feedback to counselors while they interview standardized clients (actors) in their treatment clinics. Supervisors, located remotely, listen to the interview through telephone lines and provide immediate feedback and modeling through an earpiece worn by the clinician. TCS was constructed, tested, and then adjusted based on the favorable results of an uncontrolled pilot study (Smith et al., 2007). The present study employed a randomized controlled trial to test the efficacy of workshop training plus TCS for training clinicians in MI, compared to two control conditions, workshop training followed by tape review supervision (Tape), and workshop training only (Workshop).

Method

Design overview

This study was approved by the Institutional Review Board of the NYS Psychiatric Institute as well as the Institutional Review Boards of every participating community treatment program. All participating counselors and all participating clients gave written informed consent. Figure 1 presents an overview of the recruitment, training, and assessment points over the course of the study. Directors of the participating community clinics notified counselors that there would be an opportunity to participate in an MI training study. Research staff made group presentations to the clinic staff at on-site staff meetings. The presentations explained study obligations, potential remuneration, and the inclusion/exclusion criteria. Counselors who were interested in participating were then asked to approach research staff, who conducted a brief interview to confirm eligibility and obtained written informed consent. Counselors submitted an audiotape of a counseling session with a client enrolled in their treatment clinic prior to beginning the workshop training. Clients, while enrolled in the treatment clinic, were consented research participants not familiar to the counselors (see Client-participants, below). This provided a pre-training baseline assessment of MI skill. Counselors then participated in a 2-day MI training workshop. At the end of the workshop training, counselors were randomized to one of three supervision

conditions: TCS, Tape, or Workshop. Each counselor submitted an audiotape of a counseling session with a client enrolled in their treatment clinic prior to beginning the post-workshop supervision phase of the study (Figure 1, Week 0). This provided a pre-supervision baseline assessment that included any change in MI skill level acquired during the workshop training period.

Counselors assigned to TCS completed five practice counseling sessions over the course of seven weeks (Figure 1, Training Weeks 1 to 7). Each practice session included a simulated clinical interaction with an actor that was simultaneously monitored by a supervisor who provided real-time feedback using teleconferencing technology. Counselors assigned to Tape also completed five practice counseling sessions over the seven training weeks. Each practice session included a simulated clinical interaction with an actor that was audiotaped. The audiotape was then sent to a supervisor who provided feedback several days after the practice session occurred. Counselors in the Workshop training condition received no formal feedback during the seven week training period following the workshop.

Following the supervised training phase of the study, counselors were required to submit a total of two audiotapes, each recording a counseling session with a client enrolled in their treatment clinic. These sessions were not scripted thus we could not predict if the client-participants would be more or less difficult than the scripted problems used in supervision sessions. The audiotapes were submitted one week and 12 weeks after the supervision phase ended; eight and 20 weeks, respectively, after workshop training (Figure 1, Follow-up). These assessments were conducted to assess MI skill immediately after and three-months after the supervision training. In total, counselor-participants' MI skill levels were assessed four times during the study: prior to the 2-day MI workshop (pre-workshop assessment), within seven days following completion of the workshop (post-workshop), eight weeks after the workshop (first post-supervision follow-up), and 20 weeks after the workshop (second post-supervision follow-up). MI skillfulness at each time point was assessed with the Motivational Interviewing Treatment Integrity (MITI) coding system v 2.0 (Moyers, Martin, Manual, Hendrickson, & Miller, 2005).

Design features to enhance protocol adherence—Counselor-participants were paid for their time and effort in the study. They received \$50 per workshop training day, \$25 per taped clinical interview, and a \$225 bonus for completing all assessments; a possible total of \$550 for TCS and Tape and a possible total of \$425 for Workshop. Counselor-participants were also awarded continuing education credits for their participation. Client-participants received \$15 for completing the self-report measures and the clinical interview and \$10 for completing questionnaires again 12 weeks later as a follow-up; a possible total of \$25 for each client-participant. Further, the research staff provided extensive logistical support at the clinic sites, coordinating the scheduling and conduct of all supervision sessions where counselors interviewed actors and all assessment sessions where counselors interviewed clients.

Counselor-participants

Counselors from 26 substance abuse community treatment programs affiliated with the Long Island and New York Nodes of the NIDA Clinical Trials Network (CTN) were invited to participate. Potential counselor-participants had to be between 18 and 75 years of age, provide counseling services directly to clients, and be employed at least half time at their treatment facility prior to enrolling in the study. Counselors could not participate if they had attended an MI workshop in the past three months, received training to be an MI trainer, or participated in a previous MI research trial.

Client-participants

Clients who were receiving treatment at a participating community program were invited to participate in the study by research staff or clinical supervisors at each respective clinic. Participating counselors did not recruit nor did they consent any potential client participants. Clients interested in participating were interviewed by a member of the research team. Potential client-participants had to be between 18 and 70 years of age, new to the treatment program or currently enrolled in the treatment program but new to the counselor doing the interview, using alcohol or illicit substances other than nicotine at least once per week in the 30 days prior to the interview (or prior to enrollment for residential clients), able to speak and understand English, and willing to have the session taped. Potential client-participants could not enroll in the study if they were in need of immediate treatment for a severe psychiatric disorder or medical condition, or reported legal problems that could result in incarceration for more than four weeks within three months of signing the study consent.

Supervisors

Five doctoral level clinical psychologists served as MI supervisors. All supervisors completed a standard 2-day MI training workshop, conducted by our senior trainer (LT), and then submitted taped clinical interactions that were rated for MI proficiency by the trainer (LT). Supervisors had to score at the advanced proficiency level or higher on all MITI indices (see below) before supervising counselor-participants in the present study. In addition, all supervisors received training in the use of the MITI for supervision purposes. Two supervisors (AB, KMC) and our MI workshop leader (LT) participated in MITI training programs at the University of New Mexico. Three of the five supervisors completed a "Training New Trainers" training course (JLS, KMC, DL) sponsored by the Motivational Interviewing Network of Trainers (MINT).

Assessment of MI Skill Acquisition

Motivational Interviewing Treatment Integrity (MITI) coding system v 2.0 (Moyers, Martin, et al., 2005)—The MITI is a coding system developed to quantify dimensions of an MI counseling style and has demonstrated good reliability (Pierson et al., 2007). MITI ratings were based on a randomly selected 20-minute segment of a counseling session. In the MITI framework counselors are rated along two global dimensions (Empathy and Spirit) and several behavior count measures are calculated. Empathy characterizes the extent to which the therapist understands or communicates an effort to understand a client's perspective. Spirit encompasses a counseling style that supports a client's autonomy, evokes a client's reasons for change, and fosters a collaborative counseling environment. Each global rating can range from 1 (*low*) to 7 (*high*).

The MITI behavior counts capture specific types of counselor language that are consistent with an MI counseling approach. They include open-ended questions, closed-ended questions, simple reflections, complex reflections, giving information, MI adherent behavior (MIA), and MI non-adherent behavior (MINA). These counts were used to derive five MITI indices: percentage of open questions (%OQ), percentage of complex reflections (%CR), reflection to question ratio (R:Q), percentage of MI adherent behavior (%MIA), and the number of MI non-adherent statements (MINA). An open question is defined as a question that allows for a wide range of possible answers. The percentage of open questions (%OQ) represents the number of open questions asked divided by the total number of questions during the segment. Reflections are counselor statements that convey an understanding of what the client has said. Reflections can be either simple, in which the counselor adds little or no meaning to a client's statement, or complex, in which a counselor's response adds substantial meaning or emphasis to a client's statement. The use of reflections is emphasized in MI and is seen as an important part of a counselor's MI repertoire (Miller & Moyers,

2007). The percentage of complex reflections (%CR) represents the number of complex reflections divided by the total number of reflections made. The reflection to question ratio (R:Q) is the total number of reflections made by a counselor divided by the total number of questions asked. More frequent use of reflections is considered desirable (Moyers, Martin, Manual, Miller, & Ernst, 2010). Percentage of MI adherent behavior (%MIA) is the total number of MI adherent statements divided by the total of MIAs plus MINAs (see below); higher percentage rates are more desirable. MI Adherent statements supersede all other behavior categories and are not included in the calculation of the other behavior counts.

MI non-adherent statements (MINAs) are statements made by the counselor that are inconsistent with an MI counseling style (e.g., labeling, arguing, giving advice without permission). The number of MINAs was analyzed as a count. Evidence suggests that therapist behaviors inconsistent with MI can adversely impact treatment outcome (Apodoca & Longabaugh, 2009; Guame et al., 2010; Guame et al., 2009; Miller et al., 1993; Moyers & Martin, 2006). During our pilot work, it became apparent that many counselors approached their clients in a very directive, authoritarian manner and many MINAs were of this nature. Indeed many of the clinics from which we recruited participants embraced this philosophy. Training efforts therefore focused, in part, on reducing directive and confrontational language and engendering a more MI consistent Spirit. We believe that the frequency of MINA is particularly important and thus wanted to isolate the effect of our supervision conditions on this outcome directly.

MITI ratings were calculated by supervisors. Potential biases stemming from supervisors rating tapes were addressed by blinding raters to time of assessment (i.e., baseline, post-workshop, post-supervision [week 8], or follow-up [week 20]), counselor's identity, and counselor's training condition (i.e., TCS, Tape, or Workshop). Supervisors did not rate any counselors that they had supervised. Additionally, 10% of the audiotaped sessions with clients at each assessment time point were rated by two supervisors for reliability estimates. Intra-Class Coefficients were calculated to estimate the consistency of the MITI ratings between pairs of supervisors and for the Spirit and Empathy ratings raw percent agreement was also calculated.

MI Training and Supervision

Counselor-participants were recruited and trained in eight successive waves. Each wave began with counselors providing a recorded counseling session, which was then followed by the 2-day MI workshop (see below). The beginning of each successive wave of participants was separated by eight weeks.

Randomization

Counselor-participants were randomized into one of the three supervision assignments, stratified by the clinic setting (i.e., methadone maintenance, outpatient drug-free, or residential). The randomization procedure was conducted using computer generated randomly permuted blocks. A statistician otherwise independent of the research team conducted the randomizations and held the assignments until the end of the workshop training to maintain allocation concealment during the initial training period. Randomization assignments were distributed to the participating counselors at the end of the second day of each workshop.

Motivational Interviewing Workshop

The workshop was a standard MI introductory training course that totaled 13 hours of didactics and interactive learning exercises (e.g., role plays) that took place over two days and provided participants with NYS CASAC (Certified Addictions and Substance Abuse

Counselor) credit hours. The training was provided by LT, a member of the Motivational Interviewing Network of Trainers (MINT), and an experienced workshop leader who has provided numerous MI trainings to healthcare professionals in the New York City area. MI was introduced as a counseling style that aims to help resolve an individual's ambivalence about changing certain behaviors. The workshop material covered the Spirit of MI: collaboration, evocation, and autonomy of the client; the basic principles of MI: express empathy, develop discrepancy, roll with resistance, support self-efficacy; and core MI counseling skills: the use of open questions, affirmations, reflective listening (i.e., simple and complex reflections), and summary statements (Miller & Rollnick, 2002; Miller & Moyers, 2007). In addition, recognizing and reinforcing change talk was highlighted, as recent theoretical formulations of MI propose it to be an important mechanism of change (Miller & Rose, 2009) and predictive of future client behavior (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003).

Supervision Conditions

Standard clients—The two post-workshop supervision conditions (TCS and Tape) employed actors to simulate clinical sessions. Actors were used as standard clients to provide a more homogeneous training experience across counselors and increase the likelihood of maintaining a uniform training schedule (i.e., paid actors could be relied upon to attend the scheduled practice sessions at the designated date and time). Seven standard client scripts, which varied from lower to higher stages of change (contact first author), were derived from actual clients treated at several of the participating clinics. Actors received their initial training from one of the supervisors and continued to meet every four to six weeks to ensure fidelity to the clinical scripts and address issues pertaining to the implementation of the simulated clinical encounters. The standard client-actors were deployed such that no counselor interviewed the same scripted client twice over the five supervised training sessions. The same actors and scripts were employed in both TCS and Tape conditions.

Teleconferencing supervision (TCS)—TCS was developed to maximize the acquisition and maintenance of MI skills following workshop training. It employed several supervised practice sessions that occurred in a counselor's clinical context, were distributed over time, provided immediate feedback, and differentially reinforced MIA and MINA counseling styles. Supervisors listened to counselor-actor interactions through telephone equipment and provided live feedback and coaching via a telephone ear-piece worn by the counselor during the session. The practice sessions also allowed for the modeling of MI spirit and technique by the supervisor. Midway through each practice session the supervisor and counselor would take a five-minute break to discuss the overall direction of the session and highlight the presence or absence of change talk. Live feedback during the practice sessions included short statements to reinforce MI-consistent statements (e.g., "nice open question"), to label MINA styles (e.g., "too many closed questions"), provide suggestions (e.g., "try offering a summary"), and to model MI by suggesting statements at key junctures (e.g., "ask 'How do drugs fit in your life now?'" or "say 'You're not sure there is much to gain by stopping your drug use.'").

The complete TCS post-workshop program consisted of five practice interviews each scheduled a week apart. Total supervision time per week was approximately one hour (live session plus discussion of feedback immediately prior to, and after, each session).

Tape—Tape was intended to model the supervisory practices employed in academic and research settings. Counselors recorded their sessions with actors and then sent the audiotape to the supervisor for review. The post-workshop Tape supervision program consisted of five

practice interviews each scheduled a week apart. The supervision phone call lasted approximately one hour to balance supervisory time between training conditions.

Supervisors in both TCS and Tape groups rated a randomly selected 20-minute segment of each practice session using the MITI and sent the counselor a graphical representation of their performance prior to the next scheduled practice session. The graph was accompanied with an explanation of the feedback, which included a review of important MI constructs. An example of graphical feedback and explanation of feedback can be viewed at APA's PsycARTICLES web-based supplementary materials or contact the first author. The supervisor and counselor-participant discussed the material prior to the next practice interview.

Workshop—The Workshop training condition served as a second control condition, intended to model typical workshop training not followed by supervision. Participants received no supervision following the 2-day workshop training. Counselors randomized to this condition attended the same MI introductory workshop and followed the same assessment schedule as those in the TCS and Tape conditions.

Data Analysis

The effect of supervisory condition on counselor MI skill acquisition at the post-training time points (week 8 and week 20) was tested with mixed effects linear models using a GEE analytical framework for each outcome variable, using ProcGenmod in SAS. The GEE analytical framework does not require complete outcome data across time points, so that cases missing data at one but not both outcome points were able to be included in the analyses under the assumption of missing at random. MI skill was operationalized by the two global MITI scores (i.e., Spirit and Empathy) and five indices derived from the MITI behavior counts (i.e. percent open questions, percent complex reflections, reflection to question ratio, number of MI inconsistent [non-adherent] counselor behaviors [MINA], and percent MI adherent [MIA]). Consistent with previous scoring methods (Moyers, Martin, et al., 2005), MIA was calculated as a percentage. Percent MIA, was defined as $MIA / (MIA + MINA)$. However, this scoring method entailed further missing data because cases where both MIA and MINA equal zero have a zero denominator and are thus undefined. Two outcomes, MINA and reflection to question ratio, were not normally distributed and followed the Poisson distribution, with mostly low values plus a range of higher values with positive skew. For these two outcomes, the Poisson distribution with log link function was applied and the logarithm transformed baseline scores were included as covariates in the models. We chose to investigate each MI outcome with a separate model without correction for experiment-wise error, as evidence suggests they are modestly related, although conceptually distinct, entail the use of different training exercises, and have been differentially responsive to supervision techniques (Baer et al., 2009; Miller et al., 2004).

Each linear model included contrast terms that tested the effect of supervisory condition (TCS versus Workshop; Tape versus Workshop; TCS versus Tape), time (week 8 versus week 20), and treatment clinic (i.e., residential versus methadone maintenance; outpatient drug-free versus methadone maintenance). Since the primary aims of the study were to test the effectiveness of different methods of supervision, all counselors were randomized to a supervisory condition after completing the 2-day workshop. Thus, there were two baseline assessments in this study, pre-workshop and post-workshop. Each respective model included as baseline covariate the average of the pre- and post-workshop scores. For 14 cases that were missing one of the two baseline scores, the missing scores were imputed based on the other available scores and the associations between baseline values within subjects in the

sample, using the Markov chain Monte Carlo method (Schafer, 1997); this was felt to be appropriate since the baseline scores were correlated within subjects. Interactions between supervision condition, time, and baseline scores, were initially tested in the models, but none were found significant and were therefore not included in the final models. Finally, we explored the effects of supervisory condition on several binary indicator variables, derived from the MITI scores, reflecting thresholds for declaring a clinician competent in MI skill, by constructing 2×3 contingency tables and applying chi-square or Fisher's exact test as appropriate.

Results

Participants

Figure 1 outlines the flow of counselor-participants through the study. One hundred counselors approached the research staff about participating following the presentations by our research team at staff meetings; all participants met inclusion/exclusion criteria, and signed informed consent. There were other counselors at these sites who did not approach research staff after hearing the presentations, but no data are available on their numbers or characteristics. The high rate of eligibility among those who approached the research team is likely due to the fact that as part of the presentations, eligibility criteria were described. Of the 100 counselor-participants enrolled, 97 completed the workshop and were randomized, and three were not randomized. Of three who were not randomized, one participant discontinued due to illness and two withdrew due to concerns about scheduling.

Table 1 presents the characteristics of the 97 counselors who were randomized. Overall, 52% of counselors were employed in methadone maintenance clinics, 34% in drug-free outpatient treatment clinics, and 14% in residential treatment settings. Twenty-six percent of the counselors were Latino, 40% African American, 29% Caucasian, and 5% other ethnicities. Educationally, 40% had a master's degree or higher, 31% had earned a bachelor's degree, 13% reported taking some college courses, and 15% had a high school diploma or GED. Thirty-four percent of the counselors reported they were in recovery from substance abuse problems. There were no significant differences among the counselors assigned to the three MI Training conditions on any of the demographic variables recorded.

Missing data—Of the 97 randomized, 88 (91%) had MITI data at one or both outcome points (weeks 8 and 20), and at one or both baseline points (pre- and post-workshop) and were able to be included in the primary outcome analyses (see below and Table 2). Of the nine who could not be included, five were missing both outcome scores, and four were missing both baselines. Sixteen participants were missing outcomes at week 8 (7) or week 20 (9), but not both, while 14 were missing pre- (7) or post-workshop (7) baselines, but not both. Missing data was predominantly due to poor recording quality, such that tapes of the interviews were inaudible and could not be scored.

Adherence to supervision—Among the 64 randomized to one of the two supervision conditions (TCS or Tape), adherence with supervision was high. Of the 32 counselors assigned to TCS, 30 (94%) completed all five supervision sessions, compared to 26 of 32 (81%) assigned to Tape ($\chi^2(1) = 2.29, ns$). In TCS, one participant (3%) completed four sessions, and one (3%) completed no sessions. In Tape, three participants (9%) completed four sessions, two (6%) completed two or three sessions, and one (3%) completed no sessions.

Outcomes

Reliability ratings—ICCs for the MITI global scores and behavior counts were classified according to Cicchetti's (1994) guidelines for determining the utility of clinical instruments ($<.40$ *poor*; 0.40 to 0.59 *fair*; 0.60 to 0.74 *good*; 0.75 to 1.00 *excellent*). ICCs (for a fixed set of raters; Shrout & Fleiss, 1979) for the MITI behavioral indices ranged from fair to excellent: empathy 0.50; spirit 0.60; simple reflections 0.40; complex reflections 0.64; MI non-adherent (MINA) 0.72; MI adherent (MIA) 0.47; open questions 0.74; and reflection to question ratio 0.88. Percent agreement for global scores that were calculated were: Spirit: identical ratings from two blinded coders = 37%, ratings within one point from two blinded coders = 76%; Empathy: identical ratings from two blinded coders = 34%, ratings within one point from two blinded coders = 71%.

Motivational interviewing skill acquisition—The primary aim of this study was to test the utility of different supervision techniques for improving and sustaining MI skill at eight and 20 weeks after workshop training. Parameters for each of the models are shown in Table 2, including effects of baseline MI skill score, time (week 8 versus week 20), and adjusted differences between means and estimated effect sizes (Cohen's *d*) for the contrasts between supervision conditions. There were no significant effects of the type of treatment clinic, thus these variables are not shown in the Table. Figure 2 graphs the observed means for each of the MI skill (MITI) indices by supervision group across the two baseline and two follow-up assessment points. The interactions between baseline skill level, supervision condition and time (week 8 versus week 20) were tested in each model and were not significant. Thus, the interactions were removed from the models, which were then re-estimated with only the main effect terms. In sum, the comparisons between training conditions (Effects of Supervision shown in Table 2) are adjusted for baseline skill levels and time, and estimate the differences in means between supervision conditions across the two follow-up assessment points.

Except for percent open questions (see below), there were no significant differences between the week 8 and week 20 assessments (i.e., no significant main effect of time). The overall patterning of results (Figure 2) across all skill indices suggests that counselors benefited from the workshop training, as skill levels improved from pre- to post-workshop. However, for the Workshop condition, skill levels on several of the measures deteriorate back toward baseline by the week 8 and week 20 assessment points, particularly for the Spirit and Empathy scores.

There were significant main effects of baseline skill level (i.e., average of pre- and post-workshop ratings) for all MI skill indices. Greater MI skill level prior to beginning supervision predicted greater MI skill level at the week 8 and week 20 follow-up assessments. The effects of supervisory condition on each of the MITI skill measures is now described.

Global Scores of MI Skill—The global measures of Spirit and Empathy are used to characterize important aspects of an MI style (Moyers et al., 2005). Empathy ratings capture the extent to which the therapist understands or communicates an effort to understand a client's perspective. Spirit ratings reflect the extent to which a counselor practices the three elements of the Spirit of MI (Miller & Rollnick, 2002); supporting a client's autonomy, evoking a client's own reasons for change, and setting a collaborative tone during the session. Workshop training and the feedback given to counselors addressed each global index directly. For example, a counselor learning to make an affirmation could improve his/her spirit score and a counselor learning to maintain vocal intonation to provide an accurate simple reflection could improve his/her empathy score. TCS yielded significantly greater

scores on both MI Spirit and Empathy ratings at the week 8 and week 20 follow-up points relative to the Workshop (Table 2), yielding large effect sizes (Cohen's *d*), although modest absolute differences between means. Scores for Tape at weeks 8 and 20 fell in between those of TCS and Workshop (Figure 2). Tape yielded significantly greater Spirit scores relative to Workshop, demonstrating a medium effect size, although a modest mean difference. Differences between TCS and Tape did not reach significance. As can be seen from inspection of Figure 2, all training conditions evidenced increases in Spirit and Empathy following workshop training. However, counselors in Workshop lost their gains by the week 8 assessment.

Behavior Counts of MI Skills—The goal of MI training is to lower or eliminate the occurrence of MINAs during a counseling session. Results indicated TCS yielded a significantly lower rate of MINA behavior relative to Workshop (Table 2). Again, scores for Tape fell between those of TCS and Workshop. There were no significant differences between Tape and Workshop or between TCS and Tape. Inspection of Figure 2 suggests all conditions benefitted from workshop training with MINA behaviors decreasing between the pre- and post-workshop assessments. Inspection of Figure 2 also suggests differences in the level of MINAs at the pre-workshop baseline, with higher baseline severity particularly for the TCS condition. This raises concern that differences in baseline scores could be confounding the estimates of differences in outcome between training conditions. We therefore re-ran the model, covarying only for the post-workshop baseline, where the baseline difference is less apparent. This entails a smaller sample since seven participants missing the post-workshop baseline are not included. The result is similar. The contrast between TCS and the Workshop control remains significant ($z = 2.52, p = .012$), although with a somewhat diminished *z* score, while the contrasts between Tape and Workshop, and TCS and Tape, remain non-significant.

Percent MIA is calculated as $MIA / (MIA + MINA)$. For cases where both MIA and MINA equal zero, the denominator is zero and the expression is undefined. The undefined cases had to be dropped from the analysis which yielded a smaller sample size ($n = 54$) for the model. TCS yielded significantly greater scores on percent MIA relative to Workshop. Scores for Tape again fell between those of TCS and Workshop. The difference between Tape and Workshop neared significance, while the difference between TCS and Tape was not significant.

Open questions and reflections are important component skills of MI. The goal of training is to increase their use. No statistically significant differences between supervisory conditions were detected for percent of open questions across the week 8 and week 20 assessments. Inspection of Figure 2 suggests all groups increased the use of open questions following workshop training. However, there was a significant effect of time, such that for all supervisory conditions the use of open questions declined between the week 8 and week 20 assessments. Inspection of Figure 2 again suggests all groups increased the reflection to question ratio following workshop training. TCS produced a significantly greater reflection to question ratio compared to both Workshop and Tape at the week 8 and week 20 assessments. Tape and Workshop did not significantly differ. In contrast, for the percentage of complex reflections, an opposite pattern was detected. Tape produced a significantly greater percentage of complex reflections compared to TCS, although neither Tape nor TCS differed significantly from Workshop.

Proficiency Thresholds for MI Skill—The proportion of counselors meeting each of the four proficiency thresholds for each training condition is displayed in Table 3. The thresholds were defined to reflect thresholds of competence in MI used in prior studies, but erring on the side of stringency. For example, we required that the threshold be met at both

weeks 8 and 20, thus reflecting sustained competence across time, as it was felt this would be most clinically meaningful. The threshold of global MITI scores (Spirit and Empathy) of at least five on the seven point scale represents at least a *beginner* level of proficiency, which has been used to certify counselors as therapists for clinical trials (COMBINE Study Research Group, 2003). The threshold of global scores (Spirit and Empathy) of at least six on the seven point scale represents *competence* and sets a higher bar. The thresholds of zero MINA, and 100% MIA also represent desired endpoints for training and resemble previously employed thresholds (Moyers, Miller, & Hendrickson, 2005). As can be seen in Table 3, the scores fall in the predicted direction, supervision conditions producing greater percentages of counselors who met proficiency thresholds compared to Workshop. However, these differences only meet statistical significance for the most stringent threshold (Spirit and Empathy scores of six or seven). There is little apparent difference between the two supervision conditions (TCS versus Tape). Percentages of counselors meeting the proficiency thresholds in the two supervision conditions are modest, falling in the 50% to 60% range for the three more standard measures (global MITI scores of at least five; zero MINA; 100% MIA). For the most stringent measure (global MITI scores of six or seven) the best result is in the TCS condition at only 25%.

Discussion

TCS was developed to address the limitations of standard methods for training community-based counselors in MI, by projecting expert supervision, live to counselors in the field using principles of immediate feedback, differential reinforcement, and modeling (Smith et al., 2007). In this randomized controlled trial with a diverse sample of counselors working at community-based substance abuse treatment programs, TCS was superior to workshop training alone on global scores of MI skill (Spirit and Empathy), measured with the MITI (Moyers, Martin, et al., 2005), yielding medium to large effect sizes on these measures. TCS was also superior to Workshop in reducing MINAs (e.g., argumentation and other directive or confrontational statements), which some evidence suggests have a particularly negative impact on client outcome (Guame et al., 2010; Guame et al., 2009; Miller et al., 1993; Moyers & Martin, 2006), and increasing MI adherent behaviors and the ratio of reflections to questions. For the standard tape-based supervision condition, scores on these measures for Tape fell between those for TCS and Workshop. Tape was superior to Workshop on the global Spirit score, yielding a medium-sized effect, with trends in that direction ($p < .10$) for the scores of Empathy, MI non-adherence, MI adherence, and percentage of open questions. Overall, the findings support the importance of providing feedback and supervision after workshop training to improve counselors' MI skill (Miller & Mount, 2001; Miller et al., 2004; Miller & Rose, 2009), while suggesting the promise of the methods employed in TCS.

Other aspects of the findings temper enthusiasm for TCS and suggest further improvement in training and supervision is needed. First, TCS failed to produce statistically superior MITI scores compared to the active control condition, Tape, on most measures, albeit contrasts between active intervention conditions represent a high bar. Further, the absolute differences between the group means of TCS and Workshop seem modest; e.g. in the one half to one point range on the seven point Spirit (mean difference 0.76) and Empathy (mean difference 0.68) scales.

The analyses of the proficiency thresholds are inherently less powerful, since these are categorical outcomes, but the information has arguably more direct clinical relevance for managers of treatment systems who will want to understand the extent to which implementation of MI supervision could raise the level of proficiency among their counselors. These data (Table 3) fall in the predicted direction, but differences between groups are not significant on three of the four measures. Further, the percentage of

counselors achieving proficiency across both week 8 and week 20 follow up assessments on the global Spirit and Empathy measures (i.e., a score of five *beginner proficiency* or better), the absence of MI non-adherent behavior, or 100% MI adherent behavior, fell only in the 50% to 60% range for the two supervision conditions. For the most rigorous proficiency threshold examined, global Spirit and Empathy scores of six or seven *competence*, the groups differed significantly in the predicted direction, but the highest rate, achieved by TCS, was only 25%. Of note, of the three Tape counselors who met the competency threshold at follow-up, two of them had met proficiency at baseline. And of the six TCS counselors who met the competency threshold at follow-up, three had met proficiency at baseline. This pattern suggests TCS may hold some promise in increasing the relative rates of proficiency attainment, although further improvement in supervision methods is clearly needed to increase the absolute number of counselors who attain competency following training.

Several previous randomized trials have compared methods of training counselors in MI. The results are consistent in showing that some form of training (workshop, or supervision methods based on feedback and coaching) produces superior MI skill compared to self-study (Miller et al., 2004; Moyers et al., 2008; Martino, Ball, et al., 2011). Miller and colleagues (2004) showed that training methods based on written feedback, or coaching, or both, produced superior scores compared to workshop alone on the global MI Spirit score, measured with the MISC (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003) and on the proportion of counselors who achieved a Spirit score of five or better. Outcome did not differ significantly between the three supervisory conditions. In particular, the combination condition (feedback plus coaching) failed to produce superior scores than either feedback or coaching alone, suggesting the difficulty demonstrating differences between active supervisory conditions. The counselors who participated in this trial were relatively experienced (over 80% had at least a master's degree) and probably highly motivated, having responded to advertisements and traveled to a central location to participate in the training trial. Moyers and colleagues (2008), with a sample of clinicians recruited from military mental health settings, did not find differences in MI skill between those who received workshop alone compared to those who received workshop plus subsequent supervision. Greater diversity of training, experience, and possibly motivation in their sample was invoked as a possible explanation for the less robust training effects.

In the present trial, differences between TCS and Workshop appear similar in magnitude to those observed in the trial of Miller and colleagues (2004). The present trial had a more diverse sample of counselors, recruited at community-based treatment programs, with only 40% of the counselors having a graduate degree of any kind. Tape in the present trial was similar to the feedback and coaching methods tested in the prior trials (Miller et al., 2004; Moyers et al., 2008), in that counselors received written feedback based on scores of MI skill from their practice sessions, plus the opportunity to discuss MI, the feedback, and role-play with the supervisor. TCS counselors also received written feedback and had the opportunity to interact with a supervisor. However, the focus of the supervision in TCS was on providing immediate feedback during a live interview with an actor/standard patient. The feedback was directed towards reinforcing and shaping basic component skills of MI (open questions, reflections, and eliminating MI non-adherent statements). In addition, the supervisors had the opportunity to model skills like open questions and reflections by feeding counselors lines at opportune junctures. It is curious, then, that the most robust effects of TCS, compared to Workshop, emerge on the global Spirit and Empathy scores, with more modest and inconsistent effects across the component skills. Supervisors also strove to model the spirit of MI in their approach to supervision, and it is possible that this is what came across most strongly to counselor-participants.

In addition, there is the finding, contrary to expectations, that TCS was inferior to Tape on the percentage of complex reflections. Caution is warranted in interpreting this finding, both due to limited reliability of the reflection scores, and the fact that neither Tape nor TCS was superior to Workshop on this measure. Nonetheless, it may hold another clue towards future improvements to supervision methods. Among the skills measured, the complex reflection is the most subtle and strategic, representing a culmination and tactical synthesis of a series of questions and simple reflections. It is thought to be important to the mechanism by which MI moves clients toward change (Miller & Rollnick, 2002; Miller & Rose, 2009). Tape-based supervision may have afforded more time for supervisor and counselor to discuss and role-play MI tactics and strategy outside the context of an ongoing practice interview. TCS, in its focus on the basic component skills, might have actually hindered counselors' learning of the strategic aspects of MI, perhaps by fostering over-reliance on the supervisor. Counselors' counseling behavior may end up coming too much under the control of the supervisor, rather than the cues the clinician is hearing from the client. This suggests that a more optimal form of live supervision might focus initially on shaping basic skills with "in the ear" immediate feedback, but then dialing this back, letting the counselor conduct the interview with less moment to moment intervention from the supervisor. Instead, more "time-outs" could then be taken, pausing the interview to discuss and think through the strategic issues that are unfolding. For example, rather than a direct suggestion "in the ear", such as "try a complex reflection" or "try saying this:...", the supervisor (or counselor-participant) calls a time-out, where the questioning is socratic in nature ("what's happening in the interview right now?"; "what do you want to try next?"). The effort would be to help the counselor learn to think like an MI expert, while preserving the advantage of the immediacy of the practice interview. The format of TCS with the standard patient would lend itself to frequent time-outs of this sort.

A main effect of time was observed for the outcome of percent open questions, indicating that the frequency of use of open questions, a basic component of MI, deteriorated between week 8 (immediately after the end of supervision) and the week 20 follow-up. There were no other main effects of time, nor time by training condition interactions, although inspection of the numerical trends (Figure 2), suggests some deterioration in skill between weeks 8 and 20 among the other measures. Loss of skill after workshop training alone has been fairly consistently observed across other studies on training of MI (Miller & Mount, 2001; Miller et al., 2004; Moyers et al., 2008), indicating the inadequacy of workshop training alone, and the need for post-workshop feedback and supervision. Miller and colleagues (2004) did not observe a deterioration in MI skill over long term follow-up among those counselors who received post-workshop feedback or coaching. This might, again, partly reflect the high level of prior experience and training of their sample. It would not be surprising if, for a complicated behavioral intervention such as MI, further training may be needed beyond the handful of supervision sessions offered in this and other studies (Miller et al., 2004; Moyers et al., 2008), to either sustain skills initially established at a workshop or to continue to augment those skills toward even higher levels of proficiency.

For each of the component MI skills measured, the level of counselors' skills at baseline was strongly predictive of the level of skill at the post-supervision follow-up assessments. Similar associations have been observed in previous MI training studies (Baer et al., 2004; Moyers et al., 2008). Further, no interactions between baseline skill and supervision method were detected. These results are encouraging in demonstrating that counselors across the full spectrum of MI skillfulness at baseline can benefit from a formal program of training and supervision. However, further training does not necessarily level the counseling field. Relative differences that exist prior to training are to some degree present after training. The data support the concept that active supervision may prevent loss of skill following workshop training (Table 3). Thus, counselors who benefit from workshop training may not

necessarily enhance their skill with extended supervision. However, they may be less likely to lose their newly acquired skills compared to counselors who receive no supervision.

The consistency in skill level pre- and post-supervision has a practical implication for training therapists. On the one hand, selection of therapists for training might be guided by baseline levels of proficiency, assuming the goal is to arrive at the most proficient possible pool of therapists. Given the variety of evidence-based practices available to treatment programs, as well as the limited training resources, counselors demonstrating higher baseline MI proficiency may benefit more from MI supervision than other counselors who may be better suited to providing other types of therapy. On the other hand, from the standpoint of the larger treatment system, one would not want to exclude counselors from training because of a lack of baseline skill. Again, more prolonged and improved training methods may be needed to bring the majority of counselors to high levels of performance. Martino, Canning-Ball, and colleagues (2011) have piloted an MI training method that targets poor performers for extra training, and more research on such tailored supervision methods is also indicated. Further investigation of counselor characteristics that can account for differences in MI skillfulness and response to training may be helpful in the development of tailored training interventions such as stepped training.

The present findings should be interpreted in light of both the strengths and weaknesses of the design. Strengths include: 1) high retention rates; 2) a diverse sample of counselor-participants with a wider range of educational level and experience than samples in previous training studies (Miller et al., 2004; Moyers et al., 2008); 3) randomized design with two control groups (one intended to reflect workshop training efforts typical in the field, and one representing standard supervisory techniques); and 4) blinded ratings of the primary outcome (MI skill, based on counselors' interviews with treatment-enrolled clients, using a standard measure (MITI) that demonstrated at least fair reliability on most scores). Counselors also did not select which clients to interview, another strength compared to studies where counselors could select which taped sessions to submit for review (Miller et al., 2004; Moyers et al., 2008).

The high rate of adherence represents a strength in terms of internal validity, but also raises issues of sustainability in the real world. Variable adherence was observed in a previous study of tape-based MI supervision (Miller et al., 2004). Counselors in that study often did not make tapes of their interviews, and/or were lost to follow-up. In the present study participants were paid for their time, earned continuing education credits, interacted with a motivated team of research coordinators and supervisors, and could do most of the training at their clinical site with a research coordinator handling the scheduling. Supervision in TCS, and in the other training conditions, was carried out in the spirit of MI, with an emphasis on a collaborative supervisory stance, and on positive verbal reinforcement as a primary strategy for shaping MI skills. Informal feedback from participants about their experiences across training conditions was almost uniformly positive, suggesting that training was enjoyable and rewarding. The use of actors playing standard clients surmounted logistical barriers surrounding relying on counselors to schedule clients for taping, and relying on clients to show up when scheduled (Smith et al., 2007). The scripts developed for the standard clients allowed the difficulty level and stage of change to be titrated as supervision progressed, whereas real clients would vary greatly in their presentation. The downside is that many of these features involve cost and logistics (teams of expert supervisors, trained actors, staff to coordinate scheduling, and time for individual supervision) that would be barriers to widespread dissemination of this training method into community-based treatment, where budgets and time are limited. Future research should consider modifications to TCS to reduce costs and increase efficiency, such as the development of group supervision models that might take place during regularly scheduled

staff meetings, or training individuals to serve as supervisors located in their respective clinics (Martino, Ball, et al., 2011).

There were other limitations as well: 1) Power was limited to the detection of medium to large main effects of training condition and is particularly limited to detect the interactions of treatment with baseline and time. As noted above, detecting differences in training outcome between active post-workshop supervision conditions (e.g., TCS versus Tape) is also a challenge (Miller et al., 2004) and larger samples may be needed to detect such differences; 2) The primary analyses on the seven MITI scores (Table 2 and Figure 2) were planned a priori without correction for multiple comparisons, resulting in a potentially larger study-wise error rate; 3) Raters, although blinded as noted above, were not independent of the study; 4) Client change talk has been shown to be an important predictor of client outcomes and a potential mediator of the effects of MI (Amrhein et al., 2003; Miller & Rose, 2009). Coding client change talk within the DARN-C framework (Amrhein et al., 2003), is ongoing and will be presented in a subsequent paper. Since the analyses are extensive, they are beyond the scope of this paper; 5) Counselors in the Workshop condition received less compensation than those in the supervision conditions, however, all counselor-participants received fairly generous compensation and there did not seem to be differences in adherence across conditions; 6) Reliability was in the fair to good range for most of the MITI measures, consistent with reliabilities reported in other studies (Miller et al., 2004; Moyers et al., 2008). While fair to good reliability is adequate for research purposes, consistently excellent reliability would be preferable. Further, reliability for simple reflections was weak, perhaps reflecting difficulty parsing simple from complex reflections. Limited reliability will also have an adverse impact on supervision methods, such as TCS and tape-based supervision, that are dependent on accurate ratings for the feedback provided to counselors on their performance. Thus, more work is needed on improving reliability of measurement of MI skill, both to improve research precision, and as part of a larger effort to ensure the fidelity of supervision delivered. Improving reliability would be of particular importance if supervision methods such as TCS are to be replicated and disseminated.

For psychotherapeutic interventions, the mechanisms of therapeutic effect are usually not entirely clear (Morgenstern & McKay, 2007). Components thought to be important to the effectiveness of MI (Miller & Rollnick, 2002; Miller & Rose, 2009) include the collaborative, client-centered stance, the overall guiding principles (e.g., building discrepancy, roll with resistance), and the use of counseling techniques, especially complex reflections, as well as reinforcing “change talk” or clients' statements indicating a commitment to change (Amrhein et al., 2003). Miller and Rose (2009) dichotomize critical elements of MI down to “relational variables” of empathy and spirit, and “technical variables” of therapist use of MI consistent methods. They propose a model linking training in MI, the development of specific MI skills by counselors, in-session client responses, and ultimately improved clinical outcome of clients. Studies testing training methods have the potential to refine this model to the extent that different training methods target and foster different skills. Better pinpointing of therapeutic mechanisms would in turn inform improved training methods.

In summary, this study sought to advance methods for training community-based counselors in the skill of Motivational interviewing, by testing a new supervision method, TCS, based on live supervision and immediate feedback delivered remotely by expert supervisors over the telephone. Comparison conditions were standard supervision based on review of audiotaped sessions, and a condition that received only Workshop training. The findings generally replicate the results of previous studies demonstrating the value of supervision methods involving feedback and coaching for sustaining and fostering MI skill after MI Workshop training (Miller et al., 2004). The widespread dissemination of evidence-based

practices such as MI is important for improving the quality of care delivered in the community-based treatment system. In the fields of mental health and addiction treatment, most post-graduate training in MI specifically, and new behavioral therapy techniques more generally, takes the form of courses or workshops (Walters et al., 2005; Beidas & Kendall, 2010). The present findings reinforce the recommendation that training efforts be re-oriented toward following didactic or workshop training with ongoing feedback and supervision (Grol, 2001; Miller et al., 2004; Miller & Rose, 2009).

These findings also suggest the promise of TCS, based on its robust effect on the global relational measures of MI skill (Spirit and Empathy) and on some of the measures of specific MI skills (e.g., MIA, MINA). However, weaknesses were also revealed, including poor performance of TCS on increasing complex reflections, modest impacts of both TCS and Tape in terms of raw mean differences compared to Workshop, and the proportion of counselors achieving standard thresholds of proficiency. Thus, more research is needed.

The present findings suggest a longer duration of supervision, involving a combination of techniques embodied in TCS and Tape, might be tested with the goal of bringing a greater proportion of community-based counselors to high levels of proficiency. Further, TCS and Tape, as delivered in this trial in a one-on-one format with expert supervisors, were labor intensive. More attention is also needed to modifying methods such as TCS to minimize the costs and burden, so that supervision can be brought to scale and sustained broadly across the treatment system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants from the National Institute on Drug Abuse (DA 16950; PI: E.V. Nunes), K23 DA021850 (Carpenter), K24DA022412 (Nunes).

The authors would like to thank Dr. Huiping Jiang for his outstanding contributions in the analyses of this data set.

References

- Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L. Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology*. 2003; 71:862–878. doi:10.1037/0022-006X.71.5.862. [PubMed: 14516235]
- Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction*. 2009; 104:705–715. doi:10.1111/j.1360-0443.2009.02527.x. [PubMed: 19413785]
- Baer JS, Rosengren DB, Dunn CW, Wells EA, Ogle RL, Hartzler B. An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*. 2004; 73:99–106. doi:10.1016/j.drugalcdep.2003.10.001. [PubMed: 14687964]
- Baer JS, Wells EA, Rosengren DB, Hartzler B, Beadnell B, Dunn C. Agency context and tailored training in technology transfer: A pilot evaluation of motivational interviewing training for community counselors. *Journal of Substance Abuse Treatment*. 2009; 37:191–202. doi:10.1016/j.jsat.2009.01.003. [PubMed: 19339139]
- Balcazar F, Hopkins BL, Suarez Y. A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*. 1986; 7:65–89. doi:10.1300/J075v07n03_05.
- Beidas RS, Kendall PC. Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology*. 2010; 17:1–30.

- Cicchetti DV. Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:282–290. doi: 10.1037/1040-3590.6.4.284.
- COMBINE Study Research Group. Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: Rationale and methods. *Alcoholism: Clinical and Experimental Research*. 2003; 27:1107–1122. doi:10.1111/j.1530-0277.2003.tb02873.x.
- Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*. 1999; 84:795–805. doi: 10.1037/0021-9010.84.5.795.
- Grol R. Improving the quality of medical care: Building bridges among professional pride, payer profit, and patient satisfaction. *Journal of the American Medical Association*. 2001; 286:2578–2585. doi:10.1001/jama.286.20.2578. [PubMed: 11722272]
- Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Research Into Practice I. The Lancet*. 2003; 362:1225–1230. doi:10.1016/S0140-6736(03)14546-1.
- Guame J, Bertholet N, Faouzi M, Gmel G, Daeppen JB. Counselor motivational interviewing skills and young adult change talk articulation during brief motivational interventions. *Journal of Substance Abuse Treatment*. 2010; 39:272–281. doi:10.1016/j.jsat.2010.06.010. [PubMed: 20708900]
- Guame J, Gmel G, Faouzi M, Daeppen JB. Counselor skill influences outcomes of brief motivational interventions. *Journal of Substance Abuse Treatment*. 2009; 37:151–159. doi:10.1016/j.jsat.2008.12.001. [PubMed: 19339147]
- Institute of Medicine. Board on Health Care Services. National Academy Press; Washington DC: 1998. Statement on quality of care: National Roundtable on Health Care Quality- The urgent need to improve health care quality.
- Kivlighan DM Jr, Angelone EO, Swafford. Live supervision in individual psychotherapy: Effects on therapists's intention use and client's evaluation of session and working alliance. *Professional Psychology, Research and Practice*. 1991; 22:489–495. doi:10.1037/0735-7028.22.6.489.
- Madison MB, Loignon AC, Lane C. Training in motivational interviewing: A systematic review. *Journal of Substance Abuse Treatment*. 2009; 36:101–109. doi:10.1016/j.jsat.2008.05.005. [PubMed: 18657936]
- 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. doi:10.1111/j.1360-0443.2010.03135.x. [PubMed: 20925684]
- Martino S, Canning-Ball M, Carroll KM, Rounsaville BJ. A criterion-based stepwise approach for training counselors in motivational interviewing. *Journal of Substance Abuse Treatment*. 2011; 40:357–365. doi:10.1016/j.jsat.2010.12.004. [PubMed: 21316183]
- Miller WR, Benefield G, Tonigan JS. Enhancing motivation for change in problem drinking: A controlled comparison of the two therapist styles. *Journal of Consulting and Clinical Psychology*. 1993; 61:455–461. doi:10.1037/0022-006X.61.3.455. [PubMed: 8326047]
- Miller WR, Mount KA. A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*. 2001; 29:457–471. doi:10.1017/S1352465801004064.
- Miller WR, Moyers TB. Eight stages in learning motivational interviewing. *Journal of the Teaching of the Addictions*. 2007; 5:3–14. doi:10.1300/J188v05n01_02.
- Miller, WR.; Rollnick, S. *Motivational interviewing: preparing people for change*. 2nd ed.. Guilford Press; New York, NY: 2002.
- Miller WR, Rollnick S. Ten things that motivational interviewing is not. *Behavioural and Cognitive Psychotherapy*. 2009; 37:129–140. doi:10.1017/S1352465809005128. [PubMed: 19364414]
- Miller WR, Rose GS. Toward a theory of motivational interviewing. *American Psychologist*. 2009; 64:527–537. doi:10.1037/a0016830. [PubMed: 19739882]
- Miller WR, Yahne CE, Moyers TB, 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. doi:10.1037/0022-006X.72.6.1050. [PubMed: 15612851]

- Morgernstern J, McKay JR. Rethinking the paradigms that inform behavioral treatment research for substance use disorders. *Addiction*. 2007; 102:1377–1389. doi:10.1111/j.1360-0443.2007.01882.x. [PubMed: 17610541]
- Moyers TB, Manual JK, Wilson PG, Hendrickson SML, Talcott W, Durand P. A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*. 2008; 36:149–162. doi:10.1017/S1352465807004055.
- Moyers TB, Martin T. Therapist influence on client language during motivational interviewing sessions. *Journal of Substance Abuse Treatment*. 2006; 30:245–251. doi:10.1016/j.jsat.2005.12.003. [PubMed: 16616169]
- Moyers TB, Martin T, Catley D, Harris KJ, Ahluwalia JS. Assessing the integrity of motivational interviewing interventions: Reliability of the Motivational Interviewing Skills Code. *Behavioural and Cognitive Psychotherapy*. 2003; 31:177–184. doi:10.1017/S1352465803002054.
- Moyers TB, Martin T, Houck JM, Christopher RJ, Tonigan JS. From in-session behaviors to drinking outcomes: A causal chain for motivational interviewing. *Journal of Consulting and Clinical Psychology*. 2009; 77:1113–1124. doi:10.1037/a0017189. [PubMed: 19968387]
- Moyers TB, Martin T, Manual JK, Hendrickson SML, Miller WR. Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*. 2005; 28:19–26. doi:10.1016/j.jsat.2004.11.001. [PubMed: 15723728]
- Moyers, TB.; Martin, T.; Manual, JK.; Miller, WR.; Ernst, D. Revised Global Scales: Motivational Interviewing Treatment Integrity 3.1.1 (MITI 3.1.1). 2010. <http://casaa.unm.edu/codinginst.html>. Accessed April 13, 2011
- Moyers TB, Miller WR, Hendrickson ML. How Does Motivational Interviewing Work? Therapist Interpersonal Skill Predicts Client Involvement within Motivational Interviewing Sessions. *Journal of Consulting and Clinical Psychology*. 2005; 73:590–598. doi:10.1037/0022-006X.73.4.590. [PubMed: 16173846]
- Pierson HM, Hayes SC, Gifford EV, Roget N, Padilla M, Bissett R, Fisher G. An examination of the Motivational Interviewing Integrity code. *Journal of Substance Abuse Treatment*. 2007; 32:11–17. doi:10.1016/j.jsat.2006.07.001. [PubMed: 17175394]
- Prescott P, Opheim A, Bortveit T. The effect of workshops and training on counseling skills. *Tidsskrift for Norsk Psykologforening*. 2002; 39:426–431.
- Schafer, JL. *Analysis of Incomplete Multivariate Data*. Chapman and Hall; New York, NY: 1997.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86:420–428. doi:10.1037/0033-2909.86.2.420. [PubMed: 18839484]
- Skinner, BF. *Science and Human Behavior*. The Free Press; New York, NY: 1953.
- Smith JL, Amrhein PC, Brooks A, Carpenter KM, Levin D, Schreiber EA, Nunes EV. Providing Live Supervision via Teleconferencing Improves Acquisition of Motivational Interviewing Skills After Workshop Attendance. *The American Journal of Drug and Alcohol Abuse*. 2007; 33:163–168. doi:10.1080/00952990601091150. [PubMed: 17366257]
- Walters ST, Matson SA, Baer JS, Ziedonis DM. Effectiveness of workshop training for psychosocial addiction treatments: A systematic review. *Journal of Substance Abuse Treatment*. 2005; 29:283–293. doi:10.1016/j.jsat.2005.08.006. [PubMed: 16311181]

Figure 1. Flow diagram outlining the enrollment of the 100 counselor-participants, the baseline assessments with treatment enrolled clients, the workshop training, the five supervised practice sessions with actors, and the follow-up assessments with treatment enrolled clients for each of the three supervision conditions.

Figure 2.

MITI scores (means) across time, reflecting MI skill of counselor (N = 94) randomized to supervision conditions: Teleconferencing Supervision (TCS; dashed line), Taped-based supervision (Tape; dotted line), or Workshop Only (WKSHP; solid line). Time is in weeks where time 1 (week 1) is before workshop training; time 2 (week 2) is immediately after workshop; time 8 (week 8) is after five supervision sessions, and time 20 is the week 20 follow-up.

Table 1

Demographic characteristics of the counselors (N = 97) randomly assigned to one of three conditions for training in motivational interviewing: Teleconferencing supervision (TCS), Tape-based supervision (Tape), and Workshop only (Workshop).

Variables	TCS (n=32)	Tape (n=32)	Workshop (n=33)
<u>Gender</u>			
Male	41% (n=13)	34% (n=11)	30% (n=10)
Female	59% (n=19)	66% (n=21)	70% (n=23)
<u>Age</u>	45.9 (9.5)	42.9 (11.7)	42.9 (12.4)
<u>Race</u>			
African American	28% (n=9)	44% (n=14)	48% (n=16)
Caucasian	34% (n=11)	25% (n=8)	27% (n=9)
Hispanic	38% (n=12)	19% (n=6)	22% (n=7)
Other	0% (n=0)	12% (n=4)	3% (n=1)
<u>Education</u>			
High School	12% (n=4)	9% (n=3)	24% (n=8)
College Level	44% (n=14)	41% (n=13)	45% (n=15)
Graduate Degree	38% (n=12)	47% (n=15)	33% (n=11)
<u>Treatment Clinic</u>			
Drug-Free Outpatient	41% (n=13)	31% (n=10)	30% (n=10)
Methadone Maintenance	47% (n=15)	50% (n=16)	58% (n=19)
Inpatient Residential	12% (n=4)	19% (n=6)	12% (n=4)
<u>Years in the Field</u>	8.4 (6.3)	8.5 (8.1)	7.8 (6.3)
<u>Years in current position</u>	5.5 (5.3)	4.3 (4.6)	5.7 (5.5)
<u>History of Drug Dependence</u>	34% (n=11)	41% (n=13)	27% (n=9)

Table 2

Parameter estimates, test-statistics, and estimated effect sizes for Mixed Effect Linear models testing the effect of supervision on motivational interviewing skill acquisition among counselors randomly assigned to Teleconferencing supervision (TCS), Tape-based supervision (Tape), and Workshop only (Workshop). Parameter estimates shown are regression coefficients for the baseline and time variables, and, for the effects of supervision, differences between adjusted group means.

MI Skill Indices	Parameter Estimates		
	Estimate [95% CI]	z-test Statistic	Effect Size ^a [95% CI] ^d
<u>Spirit</u>			
Baseline Spirit	0.45 [0.24; 0.66]	4.15; $p < .0001$	
Time (Week 8 vs Week 20)	0.08 [-0.21; 0.37]	0.56; $p < 0.58$	
<u>Effects of Supervision</u>			
TCS vs Workshop	0.76 [0.40; 1.12]	4.16; $p < .0001$	1.01 [0.48; 1.55]
TCS vs Tape	0.36 [-0.07; 0.79]	1.63; $p < .11$	0.48 [-0.05; 1.01]
Tape vs Workshop	0.40 [0.00; 0.80]	1.97; $p < .05$	0.53 [0.01; 1.05]
<u>Empathy</u>			
Baseline Empathy	0.41 [0.21; 0.62]	4.01; $p < .0001$	
Time (Week 8 vs Week 20)	0.06 [-0.20; 0.32]	0.45; $p < .66$	
<u>Effects of Supervision</u>			
TCS vs Workshop	0.68 [0.28; 1.09]	3.33; $p < .001$	0.74 [0.22; 1.26]
TCS vs Tape	0.34 [-0.08; 0.76]	1.57; $p < .12$	0.37 [-0.17; 0.89]
Tape vs Workshop	0.35 [-0.01; 0.71]	1.88; $p < .06$	0.38 [-0.14; 0.90]
<u>MI Non-Adherent^e</u>			
Baseline MINA (logged)	3.56 [2.25; 5.63]	5.42; $p < .0001$	
Time (Week 8 vs Week 20)	0.78 [0.46; 1.34]	-0.89; $p < .38$	
<u>Effects of Supervision</u>			
TCS vs Workshop	0.33 [0.16; 0.68]	-3.03; $p < .003$	0.70 [0.18; 1.22]
TCS vs Tape	0.62 [0.27; 1.47]	-1.08; $p < .29$	0.21 [-0.31; 0.74]
Tape vs Workshop	0.53 [0.26; 1.10]	-1.70; $p < .09$	0.49 [-0.03; 1.01]
<u>Percent MI Adherent</u>			
Baseline Percent MIA	0.53 [0.34; 0.72]	5.40; $p < .0001$	
Time (Week 8 vs Week 20)	-2.09 [-13.7; 9.56]	-0.35; $p < .73$	
<u>Effects of Supervision</u>			
TCS vs Workshop	23.3 [9.69; 36.90]	3.36; $p < .001$	0.59 [0.08; 1.11]
TCS vs Tape	7.59 [-7.00; 22.17]	1.02; $p < .31$	0.19 [-0.33; 0.72]
Tape vs Workshop	15.7 [-0.32; 31.74]	1.92; $p < .06$	0.40 [-0.12; 0.92]
<u>Percent Open Questions</u>			
Baseline Open Quest.	0.56 [0.34; 0.77]	5.05; $p < .0001$	
Time (Week 8 vs Week 20)	6.76 [1.73; 11.8]	2.64; $p < .009$	
<u>Effects of Supervision</u>			
TCS vs Workshop	4.67 [-3.88; 13.2]	1.07; $p < .29$	0.20 [-0.30; 0.71]

MI Skill Indices	Parameter Estimates		
	Estimate [95% CI]	z-test Statistic	Effect Size ^a [95% CI] ^d
TCS vs Tape	-0.92 [-9.22; 7.38]	0.22; $p < .83$	-0.04 [-0.56; 0.48]
Tape vs Workshop	5.59 [-0.82; 12.0]	1.71; $p < .09$	0.24 [-0.27; 0.76]

MI Skill	Parameter Estimates		
	Estimate [95% CI]	Z-Test Statistic	Effect Size ^a [95% CI] ^b
Percent Complex Reflections			
Baseline Complex Reflections	0.43 [0.22; 0.65]	3.88; $p < .0001$	
Time (Week 8 vs Week 20)	0.17 [-7.44; 7.79]	0.04; $p < .97$	
Effects of Supervision			
TCS vs Workshop	-7.07 [-15.7; 1.58]	-1.60; $p < .11$	-0.26 [-0.77; 0.24]
TCS vs Tape	-9.68 [-18.8; -0.56]	-2.08; $p < .04$	-0.36 [-0.89; 0.17]
Tape vs Workshop	2.62 [-6.99; 12.2]	0.54; $p < .60$	0.10 [-0.42; 0.61]
Reflection to Question Ratio^c			
Baseline R:Q Ratio (logged)	2.84 [1.32; 6.10]	2.67; $p < .01$	
Time (Week 8 vs Week 20)	1.05 [0.81; 1.36]	0.37; $p < .71$	
Effects of Supervision			
TCS vs Workshop	1.46 [1.05; 2.02]	2.28; $p < .03$	0.56 [0.04; 1.07]
TCS vs Tape	1.44 [1.01; 2.07]	2.00; $p < .05$	0.54 [0.01; 1.08]
Tape vs Workshop	1.01 [0.75; 1.36]	0.07; $p < .95$	0.01 [-0.5; 0.53]

Note. Parameter estimates shown are regression coefficients for the baseline and time variables, and, for the effects of supervision, differences between adjusted group means. All baseline by training condition and time by training condition interactions were tested and failed to reach significance. Baseline skill scores are the average of the pre- and post-workshop MI ratings.

^aEffect sizes are based on comparisons between the adjusted group means. A positive sign indicates that the effect favors the training condition presented first. The standard deviation was calculated by the pre-workshop MI ratings among control group (Workshop).

^b95% confidence intervals of effect sizes are based on raw effect sizes.

^cFor two outcomes, MINA and R:Q, Poisson regression was used, and parameter estimates shown for the effects of baseline and time are adjusted odds ratios.

Table 3

Proportion of counselors randomized to one of three training conditions meeting thresholds for proficiency on Motivational Interviewing Treatment Integrity (MITI) scores across both week 8 and week 20 follow up assessments.

MITI Proficiency Threshold	Training Condition			
	Workshop Only	Tape-based Supervision	Tele-conferencing Supervision	Chi-square (<i>df</i> = 2)
Spirit and Empathy scores = 5 or better ^a	30% (8/27)	46% (11/24)	50% (12/24)	2.47, <i>p</i> = .291
Spirit and Empathy scores = 6 or better ^a	0% (0/27)	13% (3/24)	25% (6/24)	<i>p</i> = .017 ^b
MI non-adherent count = 0	37% (10/27)	67% (16/24)	65% (15/23)	5.82, <i>p</i> = .058
MI adherent ^c 100%	33% (7/21)	53% (9/17)	50% (8/16)	1.15, <i>p</i> = .417

Note. Totals in the denominators vary due to missing data.

^a Scores of 5 are considered a *beginner* level of proficiency on the MITI; scores of 6 or 7 are considered to be in the *competent* range.

^b Fisher's Exact test calculated due to low expected frequencies.

^c MI Adherent percentage = $100 \times \text{MIA count} / (\text{MIA count} + \text{MINA count})$; cases where both MIA and MINA counts equal zero are undefined, because the denominator is zero, so these are not included, contributing to missing data.