

A Randomized Trial of Methods to Help Clinicians Learn Motivational Interviewing

William R. Miller, Carolina E. Yahne, Theresa B. Moyers, James Martinez, and Matthew Pirritano
University of New Mexico

The Evaluating Methods for Motivational Enhancement Education trial evaluated methods for learning motivational interviewing (MI). Licensed substance abuse professionals ($N = 140$) were randomized to 5 training conditions: (a) clinical workshop only; (b) workshop plus practice feedback; (c) workshop plus individual coaching sessions; (d) workshop, feedback, and coaching; or (e) a waiting list control group of self-guided training. Audiotaped practice samples were analyzed at baseline, posttraining, and 4, 8, and 12 months later. Relative to controls, the 4 trained groups showed larger gains in proficiency. Coaching and/or feedback also increased posttraining proficiency. After delayed training, the waiting list group showed modest gains in proficiency. Posttraining proficiency was generally well maintained throughout follow-up. Clinician self-reports of MI skillfulness were unrelated to proficiency levels in observed practice.

There is a widely acknowledged gap between research and practice in the treatment of substance use disorders (Institute of Medicine, 1998). Because evidence-based treatments do not automatically or even readily transfer into clinical practice, it is important to understand how new interventions can be effectively conveyed to practitioners (Rogers, 1995).

The clinical trial format provides an appropriate design for studying clinical skill transfer. In the standard clinical trial, a specified treatment is delivered to a population of clients with known characteristics, and behavioral outcomes are measured to determine the impact of intervention. Although clinical training events (such as continuing education workshops) are seldom subjected to stringent evaluation, the logic of randomized trials is directly applicable. A specified intervention (training) can be delivered to a population of participants (clinicians) with known characteristics, and behavioral outcomes (practice behaviors) can be measured to determine the efficacy of training. As with treatment assessment research, the question “Does training work?” is far too simplistic. More appropriate are questions such as “What does it take to change practice behavior?” “How do various kinds of training interact with clinician attributes?” and “Is the impact on practice behavior larger or longer lasting when certain training approaches are used or added?” To rephrase the classic statement of Gordon Paul (1969) regarding behavior therapy, the challenge is

to determine what training methods, offered by whom, to whom, and focusing on what intervention approaches, will be effective in changing what practice behaviors and for how long.

Motivational Interviewing (MI)

This study focused on methods for helping substance abuse clinicians learn the clinical method of MI. For a variety of reasons, MI seemed a good choice as subject matter for a study of treatment transfer. Its clinical procedures are well specified (Miller & Rollnick, 1991, 2002; Miller, Zweben, DiClemente, & Rychtarik, 1992), and there is reasonable evidence that MI is an efficacious treatment for substance use disorders (Burke, Arkowitz, & Dunn, 2002; Dunn, DeRoo, & Rivara, 2001; Noonan & Moyers, 1997). MI has been shown in dozens of controlled trials to produce significant change in client health behaviors in general and in substance use in particular. Randomized trials showing benefit from MI have been completed with clients whose presenting problems included alcohol (e.g., Baer et al., 1992; Bien, Miller, & Boroughs, 1993; Brown & Miller, 1993; Handmaker, Miller, & Manicke, 1999; Marlatt et al., 1998; Miller, Benefield, & Tonigan, 1993; Miller, Sovereign, & Kreege, 1988), heroin (Saunders, Wilkinson, & Phillips, 1995), marijuana (Stephens, Roffman, & Curtin, 2000), dual diagnosis (e.g., Graeber, Moyers, Griffith, Guajardo, & Tonigan, 2003), gambling (Hodgins, Currie, & el-Guebaly, 2001), hypertension (Woollard et al., 1995), and weight control in Type 2 diabetes (Smith, Heckemeyer, Kratt, & Mason, 1997). MI is a relatively brief intervention, consistent with current trends in managed care and with high early dropout rates in drug abuse treatment. MI is also compatible with a range of conceptual orientations and with current practices of many clinicians. The demand for professional training in MI is high, and such training is already widespread, yet relatively little is known about optimal methods for helping clinicians to acquire and maintain proficiency in this method.

Rollnick and Miller (1995) emphasized that MI is not a set of specific techniques but a skilled style of counseling that requires careful training. Clients' outcomes after substance abuse treatment vary widely, depending on the therapist to whom they were as-

William R. Miller, Carolina E. Yahne, Theresa B. Moyers, and Matthew Pirritano, Department of Psychology, and Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico; James Martinez, Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico.

This research was supported in part by National Institute on Drug Abuse Grant R01-DA13081 and National Institute on Alcohol Abuse and Alcoholism Grant K05-AA00133. All study and consent procedures were reviewed and approved by the Human Research Institutional Review Board of the University of New Mexico.

Correspondence concerning this article should be addressed to William R. Miller, Department of Psychology, University of New Mexico, Albuquerque, NM 87131-1161. E-mail: wrmiller@unm.edu

signed (McLellan, Woody, Luborsky, & Goehl, 1988; Najavits & Weiss, 1994). Although the determinants of therapist efficacy are not well understood, it appears that accurate empathy, as operationally defined by Carl Rogers and his students, is a measurable marker of therapists whose clients show greater improvement (e.g., Miller & Baca, 1983; Miller, Taylor, & West, 1980; Valle, 1981). Accurate empathy is a learnable clinical skill (Egan, 1975; Miller, Hedrick, & Orlofsky, 1991; Truax & Carkhuff, 1967) and has been a key component of MI (Miller & Rollnick, 2002).

The Efficacy of Therapist Training

To date, most empirical research on training of behavioral interventions has focused on teaching brief intervention strategies to primary health care personnel, including physicians, nurses, and dentists (Hollis, Lichtenstein, Vogt, Stevens, & Biglan, 1993; Hollis et al., 1994; Levin, Owen, Stinchfield, Rabinowitz, & Pace, 1999; Lichtenstein et al., 1996; Lindsay et al., 1994; Lockyer et al., 1996). The period of training in these studies has usually been short (1–2 hr). Ockene and her colleagues, for example, reported outcomes of training medical personnel in brief intervention for both smoking (Ockene et al., 1988) and heavy drinking (Ockene, Wheeler, Adams, Hurley, & Hebert, 1997). In both studies, medical staff were trained in patient-centered counseling methods similar to the MI style. Training was brief (total of 3 hr) and focused on relatively circumscribed eliciting responses (primarily asking questions). On 4-point ratings of skillfulness, trainees in both studies showed an average improvement of 0.7 points during taped interactions with standard-client actors shortly after training. In the second (alcohol) study, a single 10-min individual training consultation was added to the 3-hr group training, with a slightly larger change observed (0.73 vs. 0.68 points). Although the change in practice behavior was statistically significant at the end of training, the absolute magnitude of effect was still relatively small. Modest gains in global proficiency ratings (0.7 points on a 6-point rating scale) have also been reported following 5 days of physician training (Levin et al., 1999).

In contrast, continuing professional education for substance abuse specialists often focuses on the acquisition of more complex therapeutic methods. A common format for such training is a clinical workshop of 1–2 days in duration. Practitioners regularly attend such training on clinical innovations and indeed are usually mandated to do so to maintain professional licensure (VandeCreek & Brace, 1991). The evaluation of such training is typically limited to postworkshop questionnaires completed by participants. Self-reported benefit may not accurately reflect actual change in practice behavior, just as clients' behavioral self-report can be discrepant from more objective measures of outcome (cf. Schwarz, 1999).

When anything more than participant questionnaires has been used to evaluate training, the documentation of change has most often focused on global ratings of competence and protocol adherence (e.g., Levin et al., 1999; Lockyer et al., 1996), a tradition established by Carl Rogers and his students (Truax & Carkhuff, 1967). In the National Institute on Drug Abuse Collaborative Cocaine Treatment Study, for example, Crits-Christoph et al. (1998) found improvement across four successive training cases (on global ratings) in therapists who were learning manual-guided cognitive therapy but not for those trained in 12-step drug counseling or supportive-expressive psychotherapy for cocaine dependence. Global ratings of adherence and competence often do not

differ among compared treatments and may show little change over the course of training (Rounsaville, Chevron, Weissman, Prusoff, & Frank, 1986). More specific ratings may better differentiate treatments (DiClemente, Carroll, Connors, & Kadden, 1994; Henry, Schacht, Strupp, Butler, & Binder, 1993; Henry, Strupp, Butler, Schacht, & Binder, 1993), and behavioral measures may be still more sensitive markers of therapist skillfulness and its impact on client outcomes (e.g., Miller et al., 1993).

Evaluation of Training in MI

Several studies have focused specifically on the outcomes of training in MI. In a rudimentary first effort, Rubel, Sobell, and Miller (2000) compared pretraining (3–4 weeks before the workshop) and posttraining responses of participants in a workshop conducted by Miller and Rollnick. Only 44 of 115 participants (38%) completed and returned both the pretraining and posttraining questionnaires. Most of these (91%) worked in addiction treatment. Participants showed improvement ($p < .001$) on a 15-item measure of knowledge of MI. They were also presented with three clinical vignettes and asked to write what they would say next, a procedure adapted from the Helpful Responses Questionnaire (Miller et al., 1991). In their written responses, trainees showed a significant increase in MI-consistent responses ($p < .02$) and a significant decrease in MI-inconsistent responses ($p < .001$).

Other studies have used therapist adherence ratings to document treatment fidelity in clinical trials. Bien and colleagues (Bien et al., 1993) monitored session audiotapes for the proportion of MI-consistent and MI-inconsistent therapist behaviors. Within the multisite Project MATCH trial, videotapes of motivational enhancement therapy were centrally rated for therapist skillfulness, adherence to manual guidelines, and delivery of prescribed procedures (Carroll, Kadden, Donovan, Zweben, & Rounsaville, 1994).

Moving beyond global ratings, the Motivational Interviewing Skill Code (MISC)¹ was developed to study changes in practitioners' clinical proficiency before and after training in MI and has shown sound interrater reliability and factor structure (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003). In an initial evaluation of training using MISC (Miller & Mount, 2001) 22 probation counselors participated in an evaluation of the effectiveness of a 2-day MI training workshop. Prior to training, the counselors completed self-report questionnaires and submitted an audiotaped work sample of their counseling with actual clients. Within 2 days of completing the workshop, each counselor was videotaped interviewing a standard-client actor portraying an offender on probation. Again, 3 months later, counselors submitted work samples with actual clients and completed a final set of questionnaires. The MISC provided three types of information on the basis of three separate coding passes: (a) global ratings of counselor proficiency in MI, (b) behavior counts of specific MI-consistent and MI-inconsistent responses, and (c) relative talk time by counselor and client.

On summary measures from the MISC, there were significant increases after the workshop that were well maintained at 3-month follow-up. The total number of MI-consistent therapist responses increased 47% from pretraining to follow-up samples ($p < .0001$), as did the ratio of reflections to questions (76% increase; $p <$

¹ MISC is available at <http://casaa.unm.edu>

.007), a key component of MI. There was no significant decrease, however, in the number of MI-inconsistent therapist responses, and, consequently, the ratio of MI-consistent to MI-inconsistent practice behavior showed only modest change ($p < .04$). Counselor domination of talk time also did not decrease. In summary, we managed to add to the therapists' repertoire some MI-consistent responses, without decreasing their prior level of MI-inconsistent responses or altering the therapists' dominance of sessions. The result might be expected to be somewhat confusing—a mix of MI responses with established habits.

The MISC also tracks client responses that are prognostic of treatment outcome. On summary measures of positive (change talk) and negative (resistance) behaviors, clients showed no significant change in response to counselors before versus after training. This is of concern because client speech during MI is a reliable between-group and within-client predictor of sustained behavior change (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003; Miller et al., 1993). It appears, therefore, that although we significantly changed therapists' practice behavior in certain ways, the training effect was insufficient to make a difference in client response.

It was not a wholly unexpected finding. There is little empirical reason to believe that a one-shot training workshop (albeit the usual method for continuing professional education) would be sufficient to change durable practice behavior. The question, then, is how to enhance training so that broader changes are produced, particularly reductions in MI-inconsistent therapist responses that tend to drive client resistance and thereby alter outcomes.

Enhancing Learning of a Clinical Method

What factors, then, might enhance change in clinical practice toward the adoption of effective innovations, beyond the usual didactic strategies? One of the most consistent findings in motivational psychology is that systematic feedback enhances performance (Locke & Latham, 1990). Indeed, it is a truism that learning does not occur without feedback. Ziskin (1970) likened ordinary clinical practice to learning how to play golf in a dense fog. One hits the ball and has some feeling for immediate effect, but there is no reliable information to help correct one's drive. One could labor for years on a fog-bound driving range without improving much at the game. Feedback that is specific and is compared with behavioral goals generally favors performance improvement, particularly through induction of a perceived discrepancy between status and goal (cf. Kluger & DeNisi, 1998; Locke & Latham, 1990).

Beyond feedback, it can also be helpful to have an expert coach at hand, commenting positively on effective performance and offering specific advice for improvement. Positive reinforcement for successive approximations is another well-established principle of learning. These two common-sense and empirically grounded learning aids—systematic feedback and reinforced practice—are transferable strategies to include in efforts to change practice behavior.

This evaluation of MI training was conducted as a randomized trial to study the effectiveness of a training workshop and the added value of two training enhancement procedures (feedback and coaching) in changing practice behavior relevant to MI. Proficiency in MI was evaluated from practice samples obtained

before and immediately following training and then at 4, 8, and 12 months later.

Method

Design of the Trial

The Evaluating Methods for Motivational Enhancement Education (EMMEE) study recruited licensed English-speaking health professionals who treated substance use disorders and who desired to increase their proficiency in MI. Following baseline assessment, the 140 participants were assigned at random to one of five training conditions. All groups received a copy of a therapist manual (Miller & Rollnick, 1991) and a series of seven MI training videotapes (Miller, Rollnick, & Moyers 1998).

Workshop only (W). This group completed a 2-day clinical training workshop in MI conducted by William R. Miller, Carolina E. Yahne, and Theresa B. Moyers but with no further training enhancements until 8 months after the workshop. After Month 8, they were given feedback on their practice tapes.

Workshop plus feedback (WF). The WF group completed the same workshop and then received personal feedback on their practice tapes (including baseline and posttraining) throughout the remainder of the trial. After 8 months, they also received up to six individual coaching sessions.

Workshop plus coaching (WC). The WC group completed the workshop and then participated in up to six individual coaching sessions by telephone with an expert trainer in MI during the following 4 months. They did not receive personal feedback, however, until after the 8-month follow-up.

Workshop plus feedback and coaching (WFC). This group received all of the above training interventions, including initial workshop, ongoing feedback, and up to six individual coaching sessions following the workshop.

Self-training control (STC). Finally, one group was randomly assigned to a waiting list control condition. These participants received only the therapist manual and training videotapes and were asked to proceed to learn the method on their own for a period of 6 months. Following 4-month assessment, they received all components of the WFC group. Like all other groups, they completed the baseline tape in their home community prior to randomization and completed the baseline assessment measures on site on the evening before their workshop began. No STC participants dropped out of the study; all received training as intended.

The design combined between-group (primary contrasts at 4 months) and within-group comparisons (adding new training components after the 8-month follow-up). This combined design efficiently offered two different ways to evaluate training enhancements. Between-group differences in training conditions were evaluated at Months 4 and 8. After Month 8, personal performance feedback was given to the two groups that had not previously received it (Groups W and WC). Similarly, two groups (WF and STC) received personal coaching sessions for the first time after assessment Month 8. An additive effect of these later enhancements could be detected at the Month-12 assessment, which also offered information about the longer maintenance of gains across groups. The full design is outlined in Table 1.

Participants

The target for recruitment was 140 English-speaking participants desiring training in MI. Eligibility criteria included (a) English-speaking U.S. citizens or permanent residents; (b) licensed health professionals in counseling, psychology, medicine, nursing, or social work; and (c) treating five or more clients with substance use disorders per week in individual counseling sessions. Candidates were excluded who had received more than 8 hr of prior training in MI. Participants also had to be willing and able to travel to Albuquerque, New Mexico for the initial 2-day training portion of the study and to submit four work sample audiotapes of actual client counseling sessions over the course of 1 year.

Table 1
Overview of Study Design

Variable	W	WF	WC	WFC	STC
Assessment	B	B	B	B	B
Training	Workshop	Workshop	Workshop	Workshop	
Assessment	P	P	P	P	
Training adjunct		Feedback on B & P	Coaching	Feedback & coaching	
Assessment	4 months	4 months	4 months	4 months	4 months
Training					Workshop
Assessment					P
Training adjunct		Feedback on 4 months		Feedback on 4 months	Feedback on B, P, & 4 months
Assessment	8 months	8 months	8 months	8 months	8 months
Training adjunct	Feedback on B, P, 4, & 8	Feedback on 8 months	Feedback on B, P, 4, & 8	Feedback on 8 months	Feedback on 8 months & consults
Assessment	12 months	12 months	12 months	12 months	12 months
Training adjunct	Feedback on 12 months	Feedback on 12 months	Feedback on 12 months	Feedback on 12 months	Feedback on 12 months

Note. W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; STC = self-training control (waiting list); B = baseline; P = posttraining.

Participants were recruited via national mailings to the membership of the National Association of Alcoholism and Drug Abuse Counselors, Division 50 (Addictions) of the American Psychological Association, and the American Society of Addiction Medicine, as well as numerous e-mail lists for relevant professional groups. Availability of free training through the study was also advertised via the Web sites of MI² and the University of New Mexico Center on Alcoholism, Substance Abuse, and Addictions.³

Participants in the study were trained in four waves, separated by 6 months each. Within the first three waves, participants were randomized to all five groups, and those assigned to the four workshop groups were all trained together. The remaining participants, assigned to the waiting list (STC) group were mailed the therapist manual and videotapes and scheduled for the following workshop 6 months later. In the final (fourth) wave, participants were randomized only to the four immediate workshop groups.

Incentives

All participants received the 2-day MI workshop without charge. Meals and lodging were also provided for the period of training. All received, on completion of the training workshop (or in the case of the STC group, in advance of the workshop), the manual (Miller & Rollnick, 1991) and videotapes (Miller et al., 1998). They also received expert coded personal feedback of their performance on all audiotaped work samples. The time at which feedback was provided depended on group assignment (see Table 1). Participants were paid \$40 for completion of the 4-month and 8-month assessments (including practice sample tapes) and \$50 for completion of the 12-month assessment.

Assessment Procedures

Audiotaped work samples. On completion of screening and consent, applicants submitted the first of four audiotaped work samples. These consisted of an interview with an actual client being treated by the participant for a substance use disorder, of at least a 20-min duration, ideally on or before the client's third session. Role-played interviews were not accepted. For purposes of this training efficacy study, we asked participants to send us a self-selected session to reflect their best performance. Signed consent for audiotaping was obtained by the participant from the client using a form provided by the study and retained in the client's treatment file. To maintain client anonymity, these permission forms were not provided to the investigators. Instead, the participant submitted a signed affidavit stipulating that client permission had been obtained, signed, and

properly filed. No research information was requested from or regarding clients.

On the morning after completion of the 2-day training workshop, all participants were audiotaped interviewing 1 of 10 standard-client actors to provide a consistent measure of immediate skill gains from training (cf. Levin et al., 1999; Lockyer et al., 1996). Finally, at 4, 8, and 12 months after completion of workshop training, participants were again asked to provide an audiotaped practice sample of an interview with an actual client.

Baseline clinician characteristics. Each participant also completed a baseline assessment package immediately prior to training. Participants arrived on the day before training in time to complete the assessment package of 2–3 hr under controlled conditions. These instruments were completed once only as potential predictors of MI skill level and acquisition.

Helpful Responses Questionnaire (HRQ). Miller and colleagues (Miller et al., 1991) trained 190 volunteers in accurate empathy, a key component of MI. The six-item open-response HRQ was designed to assess the extent to which participants used reflective listening in counseling. Behavioral definitions from the MISC were used to rate clinicians' written responses. Interrater reliability for total HRQ reflection scores was $r = .932$, and trainees showed significant improvement in HRQ scores after training ($p < .001$). Baseline HRQ scores were only modestly correlated ($r = .45$) with posttraining scores.

Problem history. It is sometimes asserted that a personal history of recovery enhances a counselor's ability to empathize and counsel those with substance use disorders. Manohar (1973), however, found that recovering counselors in early sobriety had greater difficulty in manifesting client-centered attributes because of overidentification with their clients. We administered the Inventory of Drug Use Consequences (Miller, Tonigan, & Longabaugh, 1995) to measure the extent to which participants had experienced lifetime drug-related problems.

Self-esteem. Miller et al. (1991) found that trainees' baseline scores on the Coopersmith (1975) Self-Esteem Inventory (SEI) were predictive ($p < .01$) of posttraining empathy scores. For this reason, the SEI was included in our baseline battery.

² The Web site for MI is www.motivationalinterview.org

³ The Web site for the University of New Mexico Center on Alcoholism, Substance Abuse, and Addictions is <http://casaa.unm.edu>

Edwards Personal Preference Schedule (EPPS; Edwards, 1953). Based on Murray's theory of need states, the EPPS has been found in several studies to identify therapist attributes associated with client drinking outcomes (Miller, 1985). In Project MATCH, different EPPS therapist attributes were predictive of better client outcomes in MI versus 12-step treatments (Project MATCH Research Group, 1998). In particular, MI therapists whose clients fared better had higher scores on need for nurturance and lower scores on need for aggression, directly paralleling the findings of Schorer (1965).

Myers-Briggs Type Indicator (MBTI; Myers & McCauley, 1985). Also used in Project MATCH, the MBTI measures four continuous bipolar attributes of interpersonal style. We specifically sought to replicate reported findings (Miller et al., 1991; Project MATCH Research Group, 1998) linking empathic ability to five counselor traits: higher self-esteem (Coopersmith, 1975), feeling preference on the MBTI, and EPPS needs for aggression (low), achievement (low), and nurturance (high).

Confidential Pretraining Questionnaire (CPQ). A brief questionnaire was also completed by participants at the time of each work sample (Miller & Mount, 2001), with self-ratings of MI proficiency and utilization.

Follow-up. One month in advance of the target date for each follow-up, participants were reminded by e-mail that a practice sample tape was due. When practice samples were overdue, the participant received repeated e-mail and telephone prompts encouraging completion. Participants completed the CPQ for each follow-up tape submitted.

MISC System

Although clinician self-report measures are of interest, our core outcome measures were derived from the five work samples provided by participants. The baseline sample, required prior to randomization, provided a starting point against which to compare MI skill acquisition. The posttraining proficiency sample was our most controlled measure, collected on site the day after training had been completed, in interaction with actors portraying standard clients. Role-played interviews with actors, however, do not fully represent natural practice (Miller & Mount, 2001), and we therefore regarded the 4-month clinical practice sample as providing the best test of proximal skill utilization relative to baseline.

The MISC was developed as part of a pilot study (Miller & Mount, 2001) in which videotapes and audiotapes were coded by six professional staff, four of whom were experienced MI trainers. A three-tiered system evolved, requiring three passes through each tape by two independent coders. The coder first listened nonstop to a continuous 20-min segment at the beginning of the interview and completed a set of global (Likert) ratings of the therapist, the client, and their interaction. These were completed before the second pass behavior coding so as not to be driven primarily by specific behavior counts. In a second pass, the coders used a third-generation adaptation of a code originally developed at the Oregon Research Institute (Chamberlain, Patterson, Reid, Kavanagh, & Forgatch, 1984) and modified by Miller et al. (1993), in which every therapist utterance and every client utterance is classified via a set of mutually exclusive behavior categories. Then, in a third pass, coders listened nonstop once more, this time depressing timer switches to indicate relative talk time by therapist and client.

Tape coders were trained and supervised by Theresa B. Moyers. To prevent coder drift and to document interrater reliability, independent raters double coded a randomly selected 20% of all audiotapes; coders did not know which of their tapes would be double checked. Throughout most of the study, coders did not know whether tapes represented pretraining or follow-up samples. This naiveté was somewhat compromised, however, by the need to code pretraining tapes promptly for the first wave of trainees. Furthermore, the use of standard-client actors clearly identified the post-training proficiency samples. After initial coding of the first wave, the time point of data collection was not obvious to coders, nor were they informed of the study design.

Six summary measures from the MISC were selected a priori as the dependent variables representing MI proficiency: (a) global rating of MI

spirit, (b) percentage of MI-consistent responses, (c) ratio of reflections to questions, (d) percentage of questions that were open questions, (e) percentage of reflections that were complex reflections, and (f) therapist's percentage of in-session talk time (Miller & Mount, 2001).

Randomization

On completion of baseline assessment, participants were assigned to training conditions by urn randomization (Stout, Wirtz, Carbonari, & Del Boca, 1994) to ensure balancing of groups on key baseline variables. The following variables were entered into the urn: gender, year of highest degree, doctoral degree, and number of substance abuse clients treated per week. Neither participants nor trainers were aware of group assignment until workshop training had been completed, with the necessary exception that those assigned to the self-training control group were required to wait 6 months for their training.

Initial Training Workshop

All four 2-day MI training workshops were conducted by William R. Miller and Carolina E. Yahne; Theresa B. Moyers also participated in two of the workshops. The outline of material and exercises was consistent across the four workshops and paralleled the structure used in a pilot study (Miller & Mount, 2001). The workshop format was divided about evenly between didactics and demonstration (50%) and direct practice of skills by trainees (50%). In the course of 2 days of training, each participant was observed and coached by the trainers.

One significant change was made in workshop training on the basis of pilot study experience: the shift to a learning-to-learn orientation. In previous training, participants had been told, at least implicitly, that they would acquire MI skills during the workshop. Pilot study data indicated that trainees indeed left the workshop believing that they had acquired MI proficiency and reporting decreased interest in further MI training. Their practice tapes, however, reflected only modest skill acquisition, which was insufficient for clinical efficacy (Miller & Mount, 2001). We therefore changed our instructional set, advising trainees that during the workshop, they would learn the MI method but would not leave the workshop already proficient in the method. We focused in particular on recognizing important client responses that mark effective MI and predict successful behavior change outcomes. Client "change talk" is, in essence, a reinforcer that the counselor is on track in practicing MI. Client "resistance," in contrast, is a signal to shift counselor response to restore consonance (Miller & Rollnick, 2002). Once counselors are attuned to these cues, their clients in essence teach them MI.

Feedback

The EMMEE study tested the efficacy of two training enhancements, the first of which was personal feedback of MI performance. Feedback was sent to participants via e-mail, or when necessary by mailed hard copy, using a standard reporting form. The clinician feedback report provided summary measures from the MISC to provide trainees with specific information about their MI performance relative to expert models and to target levels of practice proficiency. The report included the full range of scores from tape coding (see <http://casaa.unm.edu/download/misc.pdf>). Participants also received an accompanying document that explained in detail the meaning of codes. For participants assigned to the feedback condition, reports on their baseline and posttraining sessions were sent within a month after completion of the training workshop. Subsequent feedback reports from follow-up assessment points were sent as soon as tapes had been coded, usually within 4–6 weeks of receipt of the practice tape. All participants received a sample of the feedback form at the time of workshop training but were never informed which of the many measures had been chosen as the primary dependent variables.

Coaching

Coaching provided opportunities for positive reinforcement, for problem-solving consultation around difficulties encountered in the practice of MI, and for demonstrating and practicing MI skills via role-played interviewing. Participants assigned to coaching conditions were offered up to six individual telephone coaching sessions of 30 min each. A standard heuristic was followed for these sessions, first eliciting the clinician's experiences and observations in using MI. Specific problems and challenges in applying MI were identified by the clinician, followed by collaborative problem solving. Coaches always included at least one behavior rehearsal role-play in each telephone session. Finally, the coach asked for and answered remaining questions about MI. Specific MISC performance feedback from practice samples was discussed only for participants assigned to receive both feedback and coaching.

Except in two cases, all six coaching sessions were provided by the same trainer. Coaching consultations were provided by Carolina E. Yahne (58%), Theresa B. Moyers (17%), and another highly experienced MI trainer, Denise Ernst (25%). The trainer initiated these contacts, arranging appointments via e-mail or telephone. Coaching sessions began about 2 weeks after the completion of training and continued biweekly until the sixth session or the 4-month follow-up had been completed. Coaches documented each session on a standard form, recording its date, time, duration, and adherence via a procedural checklist.

Hypotheses

We expected to observe the following effects within this randomized trial.

Hypothesis 1: All groups after training would show significant gains in MI proficiency.

Hypothesis 2: Groups receiving workshop training would show significantly greater gains in MI proficiency, relative to the untrained waiting list (STC) group.

Hypothesis 3: Groups receiving training enhancement (feedback and/or coaching) would show significantly greater gains than the W group.

Hypothesis 4: Groups receiving feedback (WF and WFC) would show significantly greater gains than groups not receiving feedback (W and WC).

Hypothesis 5: Groups receiving coaching (WC and WFC) would show significantly greater gains than groups not receiving coaching (W and WF).

Results

Participants

Over the course of 2 years, we recruited the target sample of 140 participants who were trained in a sequence of four MI clinical workshops. Demographic characteristics of participants are reported by group in Table 2. Women and men were equally represented in the overall sample, with a mean age of 47.6 years (range = 29–69 years). As a whole, the group reported 7.4 years of postsecondary education (roughly equivalent to a master's degree), 14.7 years of experience in counseling, and 10.6 years of experience in treating substance use disorders. All were licensed providers; 66% at a master's level, 19% with doctoral degrees, and 15% with a bachelor's degree or less, reflecting national trends toward the requirement of a master's degree for substance abuse treatment licensure. The most strongly endorsed therapeutic orientations were cognitive-behavioral (47.5%), 12-step (26.2%), and humanistic (22.0%).

Randomization and Baseline Equivalence

Urn randomization was successful in creating equivalent groups at baseline. No significant between-group baseline differences were observed on any of the six dependent measures of MI proficiency by multivariate analysis of variance (MANOVA), Wilks's lambda, $F(24, 448) = 1.02, p = .445$, or by univariate analyses of variance (smallest observed $p = .088$). Similarly, there were no baseline differences on variables entered into the urn (smallest observed univariate $p = .06$, for gender).

Table 2
Characteristics of the Evaluating Methods for Motivational Enhancement Education Study Participants

15.2 Characteristic	W	WF	WC	WFC	STC
<i>n</i>	24	33	34	26	23
% female	33	64	40	65	48
% White/non-Hispanic	79	91	85	89	100
% had read MI book	42	34	49	40	83
% master's or more	79	88	86	96	87
% doctorate	8	21	14	0	17
Postsecondary education					
<i>M</i> (years)	6.2	7.9	7.3	7.6	7.6
<i>SD</i>	1.8	4.2	2.9	3.9	3.6
Counseling experience					
<i>M</i> (years)	12.0	15.3	15.2	14.5	16.0
<i>SD</i>	5.5	7.2	8.3	8.5	8.4
Substance abuse counseling					
<i>M</i> (years)	9.9	9.1	11.0	12.1	10.9
<i>SD</i>	5.0	6.8	6.9	8.5	6.5

Note. W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; STC = self-training control (waiting list); MI = motivational interviewing.

Attrition

All 140 participants completed posttraining proficiency samples, interviewing standard-client actors. Two of these samples proved to be unusable because of recording equipment failure or inaudible tape. For the crucial follow-up point, only 76% of participants returned a usable practice sample, despite repeated prompts. Follow-up completion further dropped to 54% at 8 months and 45% at 12 months. A study using standard-patient interviews reported similar difficulties, obtaining 66% of follow-up work samples at 3-month follow-up (Levin et al., 1999). Attrition rates by group are reported in Table 3 and did not differ significantly at 4 months, $\chi^2(3, N = 106) = 3.25, p = .355$; 8 months, $\chi^2(4, N = 76) = 7.26, p = .123$; or 12 months, $\chi^2(4, N = 63) = 3.65, p = .455$.

As a check on representativeness of follow-up samples, we compared those who did versus did not provide 4-month data on the six dependent measures obtained at baseline and at posttraining assessments. In MANOVAs, the groups did not differ significantly in MI competence at baseline, $F(6, 131) = 1.16, p = .332$, or posttraining, $F(6, 129) = 1.02, p = .417$. Baseline and posttraining MI skill levels were also similar for those who did versus did not participate in 8-month follow-up and in 12-month follow-up (smallest obtained $p = .438$). Participants who completed follow-up interviews appear to be representative of the full sample.

Reliability of the Dependent Measures

Intraclass correlations (ICCs) were calculated to assess interrater reliability for each of the six dependent measures. The ICC is a more conservative estimate of interrater reliability than Pearson's r , as it adjusts for chance agreement between raters as well as systematic differences between raters. Cicchetti (1994) has proposed categories to evaluate the usefulness of ICCs in clinical instruments: below .40 = poor, .40 to .59 = fair, .60 to .74 = good, and .75 to 1.00 = excellent. To calculate ICC values for this data set, we chose a random sample of 50 tapes from the coder with the highest number of overall tapes coded (Coder 1), and those tapes were recoded by the same second coder (Coder 2). This sample yielded the following ICCs: overall spirit = .65, percent MI consistent = .71, reflection-to-question ratio = .95, percentage of open questions = .86, and percentage of complex reflections = .51. The ICC for percentage of talk time has consistently been in the .97-.99 range and was not recalculated in this sample.

Relationships Among Dependent Measures

As would be expected, our six measures of MI proficiency were related but not redundant. The intercorrelations among dependent

measures are reported in Table 4. Patterns of interrelationship were similar at posttraining (when the counselor interviewed a standard-client actor) and at 4 months (from the counselor's session with an actual client). The clearest convergence was between the global rating of MI spirit and the behavior counts for the percentage of MI-consistent responses. Counselor talk time, which was already well within the criterion level of 45% at baseline, was least related to other indices.

It is worth noting here that counselor self-reports of proficiency were at best modestly related to actual proficiency measures. At baseline, the item "I feel proficient and able to use MI in practice" predicted global MI spirit ratings, $r(140) = .209, p = .013$; MI-consistent response ratio, $r(150) = .169, p = .047$; and the reflection-to-question ratio, $r(140) = .329, p < .001$. The item "I am a good listener" was unrelated to the reflection-to-question ratio, $r(140) = -.015, p = .863$, and inversely related to the complex reflection ratio, $r(140) = -.244, p = .004$.

Time Effects

Table 5 reports the primary outcomes for this trial: summary values for each of the five randomized groups on each of the six MISC-dependent measures of MI proficiency at baseline, posttraining, 4-, 8-, and 12-month follow-up. We first compared baseline with posttraining values to determine whether proficiency had increased after workshop training. In all five groups, large gains in proficiency were observed immediately after the MI workshop in interactions with standard-client actors using MANOVAs, $F(1, 133) = 231.05, p < .0001$. MANOVAs for each group separately also showed large gains (all $ps < .0001$).

Similar gains were evident on MANOVAs for the full sample between baseline and 4-month follow-up, $F(1, 103) = 80.63, p < .0001$; 8-month follow-up, $F(1, 75) = 99.65, p < .0001$; and 12-month follow-up, $F(1, 58) = 42.91, p < .0001$. In summary, large gains in MI proficiency were maintained throughout the follow-up samples.

A missing values analysis was conducted to account for the possible impact of missing data on outcome variables. Using the expectation maximization method, we included the six main summary variables at all time points and data from the CPQ. The imputation of missing values yielded a Little's missing completely at random $\chi^2(424) = 456.160, p = .136$, indicating that the missing data were indeed random.

Trained Versus Untrained Clinicians

The comparison of trained with untrained (waiting list) participants occurred at the 4-month follow-up point. At this point, four

Table 3
Number of Participants (and Percentage) Providing Practice Samples at Each Assessment Point

Assessment	All groups	W	WF	WC	WFC	STC
Baseline	140	24	33	34	26	23
Posttraining	138 (99)	23 (96)	33 (100)	34 (100)	26 (100)	22 (96)
4 months	106 (76)	14 (58)	23 (70)	28 (82)	18 (69)	23 (100)
8 months	76 (54)	12 (50)	16 (49)	21 (62)	10 (38)	17 (74)
12 months	63 (45)	10 (42)	15 (46)	17 (44)	8 (31)	13 (57)

Note. W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; STC = self-training control (waiting list).

Table 4
Intercorrelations Among Measures of Counselor Proficiency at 4 Months

Measure	1	2	3	4	5	6
1. Overall MI spirit rating (1-7)	.241	.657*	.182	.385*	.441*	-.027
2. % MI consistent	.493*	.152	.233	.389*	.403*	-.043
3. Reflection:question ratio	.229*	.307*	.258*	.310*	.161	-.261*
4. % open questions	.189	.254*	.258*	.212	.299*	.028
5. % complex reflections	.234*	-.034	-.019	.201	.203	-.059
6. % talk time	.049	-.246*	-.124	-.099	-.103	.140

Note. Boldface diagonal values show the correlation between posttraining and 4-month values of the same variable ($n = 105$). Values below the diagonal are from posttraining practice samples ($n = 136$), and values above the diagonal are from 4-month practice samples ($n = 105$). MI = motivational interviewing.
* $p < .01$.

groups (W, WF, WC, and WFC) had received all of the intended training components. The STC group, however, had not yet received the clinical workshop. Between baseline and 4 months, the STC group showed no significant change in their MI skills, MANOVA, $F(1, 22) = 2.49, p = .129$. The group that had

received only the workshop (W) showed marginal gains, $F(1, 11) = 6.13, p = .031$. Large training effects were evident for the remaining three groups: WF, $F(1, 22) = 20.10, p < .001$; WC, $F(1, 27) = 42.78, p < .001$; and WFC, $F(1, 17) = 35.24, p < .001$.

Table 5
Primary Measures of Counselor Proficiency in Motivational Interviewing

Measure	STC	Trained	W	WF	WC	WFC
Overall MI spirit						
Baseline	3.65 (1.30)	3.67 (1.45)	3.29 (1.49)	3.82 (1.33)	3.79 (1.61)	3.65 (1.36)
Posttraining	4.91 (1.15)	4.85 (1.27)	4.57 (1.56)	5.00 (1.27)	4.76 (1.23)	5.04 (1.02)
4 months	4.13 (1.42)	5.00 (1.28)	4.14 (1.70)	5.13 (1.14)	5.04 (1.23)	5.44 (0.86)
8 months	4.18 (1.51)	4.95 (1.32)	4.58 (1.44)	5.13 (1.26)	5.00 (1.38)	5.00 (1.25)
12 months	5.15 (1.21)	5.04 (1.26)	4.40 (1.51)	5.33 (0.90)	5.24 (1.30)	4.88 (1.36)
% MI consistent						
Baseline	89.25 (11.36)	88.65 (13.95)	83.46 (18.59)	91.12 (11.27)	86.75 (15.07)	92.89 (7.98)
Posttraining	97.90 (3.29)	95.91 (8.25)	93.76 (12.48)	95.57 (7.64)	96.29 (7.20)	97.82 (4.73)
4 months	88.36 (16.74)	96.44 (8.60)	90.45 (16.13)	98.31 (2.48)	96.12 (8.07)	99.20 (1.91)
8 months	91.55 (10.62)	97.03 (5.22)	94.38 (8.76)	97.83 (3.34)	96.77 (4.42)	99.50 (1.58)
12 months	95.42 (6.45)	96.38 (6.39)	92.87 (7.00)	98.59 (3.39)	96.03 (8.48)	97.37 (2.37)
Reflection:question ratio						
Baseline	0.76 (0.61)	0.58 (0.44)	0.49 (0.50)	0.67 (0.46)	0.58 (0.36)	0.54 (0.42)
Posttraining	1.52 (0.55)	1.44 (0.78)	1.30 (0.70)	1.57 (0.89)	1.42 (0.84)	1.45 (0.64)
4 months	0.86 (0.57)	1.81 (1.74)	1.14 (1.01)	1.99 (2.05)	2.03 (1.98)	1.74 (1.27)
8 months	1.79 (1.55)	1.55 (1.15)	1.81 (2.07)	1.42 (0.75)	1.45 (0.85)	1.62 (0.74)
12 months	3.22 (6.59)	1.56 (1.08)	0.88 (0.72)	1.68 (0.79)	1.92 (1.40)	1.43 (0.90)
% open questions						
Baseline	32.61 (16.60)	36.13 (19.23)	36.22 (19.38)	36.50 (14.79)	38.41 (20.89)	32.61 (22.21)
Posttraining	57.55 (17.76)	57.48 (16.61)	62.61 (17.31)	58.70 (14.84)	58.05 (16.38)	50.45 (17.07)
4 months	42.86 (21.96)	51.60 (24.55)	46.84 (23.63)	52.98 (26.32)	53.92 (26.47)	49.93 (20.91)
8 months	43.62 (17.09)	48.59 (19.77)	48.10 (21.02)	49.47 (21.18)	48.56 (19.94)	47.82 (18.62)
12 months	57.03 (21.99)	50.67 (22.71)	50.22 (14.26)	48.76 (28.48)	54.20 (19.23)	47.32 (28.83)
% complex reflections						
Baseline	42.19 (20.80)	45.34 (23.94)	41.98 (27.73)	48.65 (21.70)	42.63 (24.32)	47.52 (25.16)
Posttraining	62.51 (13.33)	64.19 (15.97)	70.78 (16.50)	59.00 (13.17)	62.83 (16.97)	65.65 (15.66)
4 months	48.72 (20.75)	54.85 (18.52)	49.57 (17.84)	52.69 (21.34)	54.47 (16.69)	62.36 (17.00)
8 months	47.96 (17.83)	54.52 (18.34)	51.12 (14.39)	60.99 (17.88)	52.91 (18.83)	51.66 (22.12)
12 months	57.85 (11.71)	52.13 (18.51)	36.53 (14.09)	56.79 (16.48)	57.28 (20.39)	51.96 (13.78)
% clinician talk time						
Baseline	30.12 (12.08)	33.82 (15.61)	34.52 (17.95)	29.12 (13.94)	36.88 (14.89)	35.12 (15.78)
Posttraining	38.98 (10.49)	36.18 (9.45)	33.80 (7.74)	35.85 (6.91)	37.12 (12.15)	37.54 (9.60)
4 months	36.19 (16.67)	30.63 (12.75)	31.77 (15.71)	31.80 (13.84)	29.48 (11.13)	30.03 (12.04)
8 months	29.40 (11.38)	28.48 (13.49)	25.67 (10.55)	28.90 (13.53)	30.22 (15.47)	27.50 (13.46)
12 months	31.62 (12.56)	30.97 (16.01)	35.15 (15.35)	27.93 (14.74)	30.62 (19.56)	32.21 (12.05)

Note. Values are means (and standard deviations). The posttraining sample was always collected immediately after workshop training. For the self-training control (STC [waiting list]) group only, the 4-month practice sample preceded the training workshop and therefore also preceded the posttraining sample. Trained = all four trained groups (W, WF, WC, and WFC); W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; MI = motivational interviewing.

We used multivariate analysis of covariance (MANCOVA) to compare the STC group with the four trained groups on the six dependent measures at 4 months, entering baseline values of the same variables as covariates. The contrast showed significantly larger gains in MI proficiency in the four trained groups, relative to STC, $F(6, 88) = 3.39, p < .005$. This difference was also apparent on some of the univariate analyses of covariance (ANCOVAs) for the six proficiency measures: global spirit ($p = .007$), reflection-to-question ratio ($p = .012$), percentage of open questions ($p = .260$), percent complex reflections ($p = .263$), MI-consistent response ratio ($p < .001$), and therapist talk time ($p = .06$). We further decomposed the MI-consistent response ratio into its two components and found that although training did not increase the frequency of MI-consistent counselor responses, $F(1, 100) = 2.04, p = .156$, there was a large decrease in the frequency of MI-inconsistent responses after training, $F(1, 100) = 8.41, p < .005$.

Specific Effects of Training Enhancements

Also of interest was the impact of the specific training enhancements (feedback and coaching) on training outcomes. A first step here was an omnibus MANCOVA for the six dependent measures among the four trained groups (W, WF, WC, and WFC) at 4 months, using baseline values as covariates. This test detected no significant aggregate differences among the four groups, Wilks's lambda $F(18, 187) = 0.77, p = .735$, with only one univariate difference on global MI spirit ratings, $F(3, 78) = 4.25, p = .036$.

We followed up with clinical significance analyses to determine whether the differences in training outcomes were clinically meaningful. These are reflected in Figure 1 for the two composite measures that best capture overall MI style: the global MI spirit rating and the percentage of MI-consistent responses. The clinical trial proficiency standard adopted by the COMBINE Study Research Group (2003) for global MI spirit is 5 on the 7-point rating scale. All three of the groups receiving feedback and/or coaching met this standard, on average, at 4-month and 8-month follow-up, whereas neither the waiting list (STC) nor the workshop-only (W) group did so. The groups differed significantly in the percentage of clinicians meeting or exceeding the clinical proficiency criterion at 4 months, $\chi^2(4, N = 106) = 9.62, p = .047$. The percentages of trainees meeting this proficiency criterion within each group are shown in Figure 2.

Similarly, only the enhanced training conditions met the proficiency standard of 95% MI-consistent responses at 4-month and 8-month follow-ups. In this case, however, the percentages of clinicians achieving criterion proficiency did not differ significantly, $\chi^2(4, N = 76) = 7.21, p = .125$.

Impact of Counselor Training on Client Response

Another important question is whether clinician training is sufficient to produce a change in client response. We did not evaluate client outcomes in this study but do have a good proxy in that client in-session responses during MI are prognostic of behavioral outcomes (e.g., Amrhein et al., 2003; Miller et al., 1993). We therefore analyzed changes in two key client responses: change talk and resistance (Miller & Rollnick, 2002). Client responses to counselors changed significantly in the expected direction (more

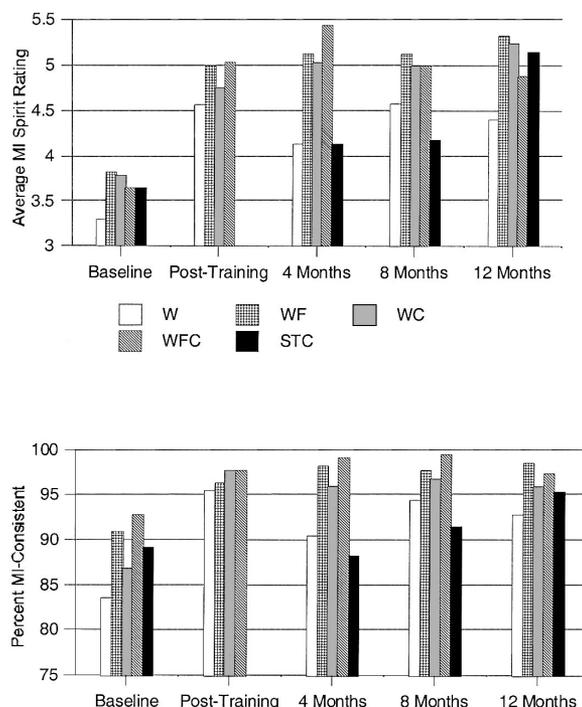


Figure 1. Changes in counselor proficiency as a function of training methods. Top: Average motivational interviewing (MI) spirit rating. Bottom: Percentage of MI-consistent responses. W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; STC = self-training control (waiting list).

change talk, less resistance) between baseline and the posttraining (with actors) proficiency sample, $F(1, 135) = 24.39, p < .0001$.

When actual client responses were measured at 4 months, however, only the WFC group showed significantly better client responses compared with baseline, $F(1, 17) = 9.97, p = .006$. The four trained groups combined thus failed to show aggregate change in client response, $F(1, 82) = 3.15, p = .080$. Here, then, we observed an advantage for the combined training methods.

Effects of Counselor Pretraining Characteristics

Finally, we examined counselor pretraining characteristics as potential predictors of ability to acquire MI proficiency. When global MI spirit score at baseline was regressed onto the five personality variables hypothesized to be predictors (self-esteem, feeling, aggression, achievement, and nurturance), no significant relationship was found ($R^2 = .037$), $F(5, 134) = 1.03, p = .405$, nor did these variables predict 4-month MI spirit score ($R^2 = .086$), $F(5, 74) = 1.39, p = .237$. Similarly, no relationship was found with counselors' MI-consistent response ratio at baseline ($R^2 = .045$), $F(5, 134) = 1.27, p = .279$, or at 4 months ($R^2 = .050$), $F(5, 74) = 0.78, p = .568$.

Counselors' personal history of drug problems did predict levels of MI proficiency, but only at baseline. Counselors with higher lifetime levels of drug problems showed significantly lower levels of MI proficiency on all six dependent measures (all $ps < .001$). After training, however, personal history no longer predicted counselor level of MI proficiency.

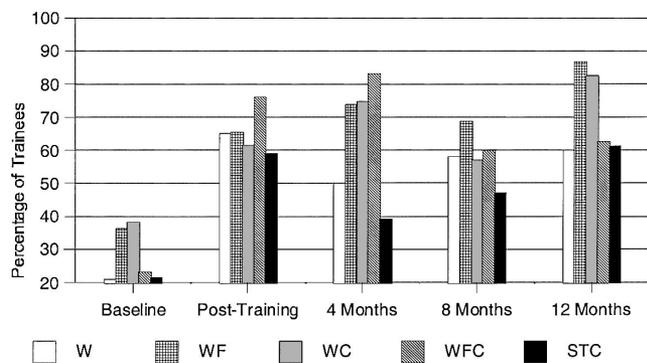


Figure 2. Clinical significance: Percentage of trainees meeting overall proficiency criterion. W = workshop only; WF = workshop plus feedback; WC = workshop plus coaching; WFC = workshop plus feedback and coaching; STC = self-training control (waiting list).

Discussion

Four Surprises

There were at least four surprises in our findings. First, on the basis of prior experience, we had not expected the 2-day training workshop to produce such large and immediate effects on MI proficiency. All four workshop groups showed a substantial jump in proficiency immediately following the workshop when interacting with a standard-client actor. This is in sharp contrast to what we had observed in a pilot study (Miller & Mount, 2001) but consistent with data recently reported by Sholomskas, Syracuse, Ball, Rounsaville, and Carroll (in press).

There are several possible reasons for this difference between studies. In the pilot study, we were training probation officers who worked together in the same program, whose supervisor had arranged the training, and who were expected to participate. In the main trial, participants were substance abuse professionals who individually enrolled for the study and training, traveled to Albuquerque at their own expense, and were compensated for participating. At baseline, EMMEE participants already showed a higher percentage of MI-consistent responses (89.3%) as compared with the pilot sample (64.4%). In the pilot study, when we offered up to six free follow-up consultation sessions on site, the median number of sessions attended was zero. In the present study, participants who were offered up to six individual follow-up telephone consultations completed five of them, on average. Four months after training, only 52% of pilot project participants reported having read the book on MI, whereas in EMMEE, the comparable figure was 87%. One reason for observed discrepancies in outcome, then, could be these differing attributes of the two samples.

As indicated earlier, we also changed our training approach between the pilot and the main trial. We placed more emphasis on the underlying spirit of MI and less on techniques. Relative to the pilot study, we observed much greater gains in our overall rating (spirit) that reflects understanding and practice of the underlying approach. Also, on the basis of the finding of little benefit from a workshop alone, we shifted to a learning-to-learn model in which we emphasized what responses to watch for in clients that signal good or not-so-good practice of MI.

A second surprise was that the W group showed an apparent reversal of these posttraining gains, lagging at 4 months near the

levels of the untrained waiting list group (see Figure 1). This group was able to demonstrate high levels of MI proficiency on the day after training, interacting with a standard-client actor, but then fell below clinical trial proficiency criteria without further training support. This reversal after one-shot workshop training is consistent, however, with a long-standing literature on staff training (Stokes & Baer, 1977). Those receiving feedback and/or coaching more fully retained clinical trial proficiency levels.

A third unexpected finding was that the STC group had not caught up with the others even 4 months after training (at Month 8). This parallels a finding from clinical trials, in which a waiting-list control group was used to evaluate the effect of treatments for depression (Schmidt & Miller, 1983) and problem drinking (Harris & Miller, 1990). In both studies, delayed treatment yielded poorer outcomes than when treatment was given immediately.

Finally, there is the finding that what actually happened in training (at least with regard to the MI-consistent response ratio) was not so much a substantial increase in MI-consistent responses (e.g., reflective listening) but rather a reduction in MI-inconsistent responses (e.g., confrontation). In a pilot study (Miller & Mount, 2001), the opposite pattern was obtained: There was a modest increase in MI-consistent responses but no decrease in confrontive MI-inconsistent responses. In essence, training in the pilot study seemed only to have sprinkled some reflective-listening salt atop a polyglot stew. Our self-selected trainees in EMMEE started with a much higher level of MI-consistent responses before training and did not really show significant increases in these responses. Instead, they showed significant increases in the overall spirit rating, and their MI-consistent response ratio improved because they decreased MI-inconsistent responding. This suggests that, as speculated by Miller and Mount (2001), it may be at least as important in teaching MI to diminish old habits of MI-incompatible responding. A similar finding was previously reported for addiction counselors' written responses to clinical vignettes following MI training (Rubel et al., 2000).

Three Caveats

An unexpected disappointment was the rapid attrition in the percentage of participants who submitted practice samples at 4, 8, and 12 months after training, a problem also reported by other investigators (Levin et al., 1999). Compliance at 4 months was high in the waiting list group (100%) because they had to submit this tape to be admitted to their (delayed) training workshop. Compliance dropped from 76% at 4 months to just over half at 8 months and a dismal 45% at 12 months. Although retained participants appear to be representative of the full sample, caution is clearly warranted in interpreting later follow-ups.

The proficiency levels shown in these practice samples probably represent a best-case picture of counselors' actual practice. They selected the client and session to tape and were well aware that it would be used to measure their proficiency in MI. One might expect, then, that they would be on their best behavior with regard to MI. The practice tapes therefore reflect ability to demonstrate MI on demand, but this is no guarantee that these same skills are used in routine practice. Monitoring in psychotherapy trials often involves random selection of therapist tapes, which was definitely not the case in this study. Evidence for practice change in this study is limited to practitioners' demonstrated facility with MI,

combined with self-report of high utilization of the method in their regular practice behind closed doors.

It is also clear that this was a self-selected sample of practitioners, apparently well motivated to learn MI. It was our intent to determine first whether training could change practice behavior under optimal conditions. Participants entered with a good head start on MI skills and established high proficiency rather quickly. The same results may not occur when training practitioners who are less interested in learning this approach are required to attend training (cf. Miller & Mount, 2001). For such clinicians, training may need more initial focus on enhancing motivation to learn this approach. The compressed range of participants may also have masked predictors of ability to acquire proficiency in MI, which would be apparent with a wider range of practitioner attributes.

Two Points of Encouragement

Within these cautions, the results of this trial are encouraging. MI is a clinical method with reasonable evidence of efficacy and for which underlying mechanisms of effect are being clarified (Amrhein et al., 2003). These findings add evidence that clinical skillfulness in MI can be acquired by practitioners who wish to learn it and within a relatively short span of months. Conversely, self-guided training using the MI book and training videotapes produced no significant increase in clinical proficiency.

It is also encouraging that we observed some changes not only in counselor behavior but also in client response. Between baseline and posttraining (with actors), client responses changed as expected: higher levels of change talk and lower levels of resistance. When examining actual client responses at 4 months after training, there was only a trend ($p < .08$) for the whole sample. Only those receiving the full training package (WFC) showed significant changes in client response ($p < .006$). These changes of in-session client response are predictive of better behavior change outcomes (Amrhein et al., 2003; Miller et al., 1993). Still better, of course, would be research documenting improvement in client outcomes after training.

Implications for Training

This study provides support for the efficacy of training in MI. Participants attending a 2-day workshop showed substantial gains in MI proficiency, as compared with no change with self-directed learning by book and videotapes. At least with this treatment approach, it appears that the dissemination of therapist manuals is insufficient to alter practice behavior.

The reversal of gains in the W group also suggests a need for "relapse prevention" measures after initial training (cf. Milne, Westerman, & Hanner, 2002). Our data indicate that coaching and/or individual feedback of performance during the months after initial training can help to maintain gains in clinical proficiency. For complex clinical skills such as MI, it is unreasonable to expect that a one-shot workshop will establish enduring competence, and ongoing support is needed for acquisition and retention of proficiency. Trainers should take this into account in planning how to help clinicians to learn MI.

It is also clear from the present study that different measures of training effects are not interchangeable. As others have found (e.g., Levin et al., 1999), self-reported clinical proficiency shares minimal variance with actual proficiency, as measured from practice

samples. We found that clinicians were able to demonstrate training gains when interviewing a standard-client actor. The difficulty of obtaining follow-up practice samples from clinicians suggests the usefulness of obtaining such a practice sample immediately after training. We did find, however, that actors' responses to clinicians were not representative of how actual clients respond to MI. The actors were relatively invariant in responding to different clinicians; they tended to reproduce the same script. Thus, if the goal is to determine how training affects client response to clinicians, standard-client actors may not be an optimal method. This also converges with our experience that when clinicians role-play clients during training, they tend to portray "clients from hell" who are far more intransigent and less responsive to MI than those encountered in actual practice.

The research described here focused on improving proficient practice one clinician at a time. When such individual practitioners return to their workplace, they may encounter little support for their newfound skills. As we contacted our participants for follow-up tapes, we found that many had changed jobs. Although high turnover is not unusual in substance abuse treatment, we wondered whether such proximal job changes might be related to training, and we began to inquire as to the reasons. Anecdotally, several participants told us that MI training had crystallized for them their dissatisfaction with the authoritarian modus operandi of the treatment system within which they had been working and that they had moved to new treatment settings more consistent with their client-centered approach. This points to the challenge of systemic rather than merely individual change in practice. The collaborative, evocative, autonomy-respecting style of MI is more compatible with some (e.g., client-centered) practice settings than with others (e.g., confrontational counseling and authoritarian therapeutic communities). One approach is to impose system change from the top down, making continued employment contingent on adoption of the desired clinical practices (e.g., Liddle et al., 2002). This is likely to inspire reactance, however, and requires continued monitoring of adherence. It is desirable for implementation to be accomplished in a manner that mirrors and models the therapeutic approach itself (Liddle et al., 2002). Here, then, is a challenge: to find model-consistent ways for disseminating MI at a systemic as well as individual practitioner level.

Summary

The EMMEE study demonstrated successful dissemination into practice of the evidence-supported clinical method of MI through systematic clinical training with real-world substance abuse practitioners. The addition of feedback and/or coaching appears to improve the retention of proficiency after a 2-day clinical workshop. Participants showed no significant improvement of proficiency through self-study using a therapist manual and training videotapes, and a 2-day workshop alone yielded modest gains in skillfulness that returned to baseline levels 4 months later. Self-study and one-time clinical workshops are common methods by which practitioners seek to acquire new clinical skills. At least with regard to MI, the efficacy of these educational methods is questionable without further support for skill acquisition and maintenance.

References

- Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology, 71*, 862–878.
- Baer, J. S., Marlatt, G. A., Kivlahan, D. R., Fromme, K., Larimer, M., & Williams, E. (1992). An experimental test of three methods of alcohol risk-reduction with young adults. *Journal of Consulting and Clinical Psychology, 60*, 974–979.
- Bien, T. H., Miller, W. R., & Borouhgs, J. M. (1993). Motivational interviewing with alcohol outpatients. *Behavioural and Cognitive Psychotherapy, 21*, 347–356.
- Brown, J. M., & Miller, W. R. (1993). Impact of motivational interviewing on participation in residential alcoholism treatment. *Psychology of Addictive Behaviors, 7*, 211–218.
- Burke, B. L., Arkowitz, H., & Dunn, C. (2002). The efficacy of motivational interviewing and its adaptations: What we know so far. In W. R. Miller & S. Rollnick (Eds.), *Motivational interviewing: Preparing people for change* (2nd ed., pp. 217–250). New York: Guilford Press.
- Carroll, K. M., Kadden, R. M., Donovan, D. M., Zweben, A., & Rounsaville, B. (1994). Implementing treatment and protecting the validity of the independent variable in treatment matching studies. *Journal of Studies on Alcohol, 12*, 149–155.
- Chamberlain, P., Patterson, G., Reid, J., Kavanagh, K., & Forgatch, M. (1984). Observation of client resistance. *Behavior Therapy, 15*, 144–155.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.
- COMBINE Study Research Group. (2003). Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: Rationale and methods. *Alcoholism: Clinical and Experimental Research, 27*, 1107–1122.
- Coopersmith, S. (1975). *Self-Esteem Inventory*. Lafayette, CA: Self-Esteem Institute.
- Crits-Christoph, P., Siqueland, L., Chittams, J., Berger, J. P., Beck, A. T., Frank, A., et al. (1998). Training in cognitive, supportive–expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology, 66*, 484–492.
- DiClemente, C. C., Carroll, K. M., Connors, G. J., & Kadden, R. M. (1994). Process assessment in treatment matching research. *Journal of Studies on Alcohol, 12*, 156–162.
- Dunn, C., DeRoo, L., & Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review. *Addiction, 96*, 1725–1742.
- Edwards, A. L. (1953). *Edwards Personal Preference Schedule*. New York: Psychological Corporation.
- Egan, G. (1975). *The skilled helper*. Monterey, CA: Brooks/Cole.
- Graeber, D. A., Moyers, T. B., Griffith, G., Guajardo, E., & Tonigan, S. (2003). A pilot study comparing motivational interviewing and an educational intervention in patients with schizophrenia and alcohol use disorders. *Community Mental Health, 39*, 189–202.
- Handmaker, N. S., Miller, W. R., & Manicke, M. (1999). Findings of a pilot study of motivational interviewing with pregnant drinkers. *Journal of Studies on Alcohol, 60*, 285–287.
- Harris, K. B., & Miller, W. R. (1990). Behavioral self-control training for problem drinkers: Components of efficacy. *Psychology of Addictive Behaviors, 4*, 82–90.
- Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists' response to training. *Journal of Consulting and Clinical Psychology, 61*, 441–447.
- Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology, 61*, 434–440.
- Hodgins, D. C., Currie, S. R., & el-Guebaly, N. (2001). Motivational enhancement and self-help treatments for problem gambling. *Journal of Consulting and Clinical Psychology, 69*, 50–57.
- Hollis, J. F., Lichtenstein, E., Vogt, T. M., Stevens, V. J., & Biglan, A. (1993). Nurse-assisted counseling for smokers in primary care. *Annals of Internal Medicine, 118*, 521–525.
- Hollis, J. F., Vogt, T. M., Stevens, V., Biglan, A., Severson, H., & Lichtenstein, E. (1994). The tobacco reduction and cancer control (TRACC) program: Team approaches to counseling in medical and dental settings. In D. M. Burns, S. J. Cohen, T. E. Kottke, & E. R. Gritz (Eds.), *Tobacco and the clinician: Intervention for medical and dental practice* (pp. 143–173). Bethesda, MD: National Cancer Institute.
- Institute of Medicine. (1998). *Bridging the gap between practice and research: Forging partnerships with community-based drug and alcohol treatment*. Washington, DC: National Academy Press.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science, 7*, 67–72.
- Levin, F. R., Owen, P., Stinchfield, R., Rabinowitz, E., & Pace, N. (1999). Use of standardized patients to evaluate the physicians in residence program: A substance abuse training approach. *Journal of Addictive Diseases, 18*, 39–50.
- Lichtenstein, E., Hollis, J. F., Severson, H. H., Stevens, V. J., Vogt, T. M., Glasgow, R. E., & Andrews, J. A. (1996). Tobacco cessation interventions in health care settings: Rationale, model, outcomes. *Addictive Behaviors, 21*, 709–720.
- Liddle, H. A., Rowe, C. L., Quille, T. J., Dakof, G. A., Mills, D. S., Sakran, E., & Biaggi, H. (2002). Transporting a research-based adolescent drug treatment into practice. *Journal of Substance Abuse Treatment, 22*, 1–13.
- Lindsay, E. A., Ockene, J. K., Berger, L., Hymowitz, N., Pomrehn, P., & Wilson, D. M. (1994). Physicians' and dentists' roles in COMMIT—The Community Intervention Trial for Smoking Cessation. In D. M. Burns, S. J. Cohen, T. E. Kottke, & E. R. Gritz (Eds.), *Tobacco and the clinician: Intervention for medical and dental practice* (pp. 334–341). Bethesda, MD: National Cancer Institute.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Lockyer, J., el-Guebaly, N., Simpson, E., Gromoff, B., Toews, J., & Juschka, B. (1996). Standardized patients as a measure of change in the ability of family physicians to detect and manage alcohol abuse. *Academic Medicine, 71*, 81–83.
- Manohar, V. (1973). Training volunteers as alcoholism treatment counselors. *Quarterly Journal of Studies on Alcohol, 34*, 869–877.
- Marlatt, G. A., Baer, J. S., Kivlahan, D. R., Dimeff, L. A., Larimer, M. E., Quigley, L. A., et al. (1998). Screening and brief intervention for high-risk college student drinkers: Results from a 2-year follow-up assessment. *Journal of Consulting and Clinical Psychology, 66*, 604–615.
- McLellan, A. T., Woody, G. E., Luborsky, L., & Goehl, L. (1988). Is the counselor an “active ingredient” in substance abuse rehabilitation? An examination of treatment success among four counselors. *Journal of Nervous and Mental Disease, 176*, 423–430.
- Miller, W. R. (1985). Motivation for treatment: A review with special emphasis on alcoholism. *Psychological Bulletin, 98*, 84–107.
- Miller, W. R., & Baca, L. M. (1983). Two-year follow-up of bibliotherapy and therapist-directed controlled drinking training for problem drinkers. *Behavior Therapy, 14*, 441–448.
- Miller, W. R., Benefield, R. G., & Tonigan, J. S. (1993). Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology, 61*, 455–461.
- Miller, W. R., Hedrick, K. E., & Orlofsky, D. (1991). The Helpful Re-

- sponses Questionnaire: A procedure for measuring therapeutic empathy. *Journal of Clinical Psychology*, 47, 444–448.
- Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*, 29, 457–471.
- Miller, W. R., & Rollnick, S. (1991). *Motivational interviewing: Preparing people to change addictive behavior*. New York: Guilford Press.
- Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). New York: Guilford Press.
- Miller, W. R., Rollnick, S., & Moyers, T. B. (1998). *Motivational interviewing* (6-tape series). Albuquerque: University of New Mexico.
- Miller, W. R., Sovereign, R. G., & Krege, B. (1988). Motivational interviewing with problem drinkers: II. The Drinker's Check-Up as a preventive intervention. *Behavioural Psychotherapy*, 16, 251–268.
- Miller, W. R., Taylor, C. A., & West, J. C. (1980). Focused versus broad spectrum behavior therapy for problem drinkers. *Journal of Consulting and Clinical Psychology*, 48, 590–601.
- Miller, W. R., Tonigan, J. S., & Longabaugh, R. (1995). *The Drinker Inventory of Consequences (DrInC): An instrument for assessing adverse consequences of alcohol use—Test manual*. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- Miller, W. R., Zweben, A., DiClemente, C. C., & Rychtarik, R. G. (1992). *Motivational enhancement therapy manual: A clinical research guide for therapists treating individuals with alcohol abuse and dependence* (Vol. 2, Project MATCH monograph series). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- Milne, D., Westerman, C., & Hanner, S. (2002). Can a “relapse prevention” model facilitate the transfer of training? *Behavioural and Cognitive Psychotherapy*, 30, 361–364.
- Moyers, T. B., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the Motivational Interviewing Skills Code. *Behavioural and Cognitive Psychotherapy*, 31, 177–184.
- Myers, I. B., & McCauley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Najavits, L. M., & Weiss, R. D. (1994). Variations in therapist effectiveness in the treatment of patients with substance use disorders: An empirical review. *Addiction*, 89, 679–688.
- Noonan, W. C., & Moyers, T. B. (1997). Motivational interviewing: A review. *Journal of Substance Misuse*, 2, 8–16.
- Ockene, J. K., Quirk, M. E., Goldberg, R. J., Kristeller, J. L., Donnelly, G., Kalan, K. L., et al. (1988). A residents' training program for the development of smoking intervention skills. *Archives of Internal Medicine*, 148, 1039–1045.
- Ockene, J. K., Wheeler, E. V., Adams, A., Hurley, T. G., & Hebert, J. (1997). Provider training for a patient-centered alcohol counseling in a primary care setting. *Archives of Internal Medicine*, 157, 2334–2341.
- Paul, G. L. (1969). Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status* (pp. 29–62). New York: McGraw-Hill.
- Project MATCH Research Group. (1998). Therapist effects in three treatments for alcohol problems. *Psychotherapy Research*, 8, 455–474.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.
- Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, 23, 325–334.
- Rounsaville, B. J., Chevron, E. S., Weissman, M. M., Prusoff, B. A., & Frank, E. (1986). Training therapists to perform interpersonal psychotherapy in clinical trials. *Comprehensive Psychiatry*, 27, 364–371.
- Rubel, E. C., Sobell, L. C., & Miller, W. R. (2000). Do continuing education workshops improve participants' skills? Effects of a motivational interviewing workshop on substance abuse counselors' skills and knowledge. *Behavior Therapist*, 23, 73–77.
- Saunders, B., Wilkinson, C., & Phillips, M. (1995). The impact of a brief motivational intervention with opiate users attending a methadone programme. *Addiction*, 90, 415–424.
- Schmidt, M. M., & Miller, W. R. (1983). Amount of therapist contact and outcome in a multidimensional depression treatment program. *Acta Psychiatrica Scandinavica*, 67, 319–332.
- Schorer, C. G. (1965). Defiance and healing. *Comprehensive Psychiatry*, 6, 184–190.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105.
- Sholomskas, D. E., Syracuse, G., Ball, S. A., Rounsaville, B. J., & Carroll, K. M. (in press). We don't train in vain: A controlled trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*.
- Smith, D. E., Heckemeyer, C. M., Kratt, P. P., & Mason, D. A. (1997). Motivational interviewing to improve adherence to a behavioral weight-control program for older obese women with NIDDM: A pilot study. *Diabetes Care*, 20, 53–54.
- Stephens, R. S., Roffman, R. A., & Curtin, L. (2000). Comparison of extended versus brief treatments for marijuana use. *Journal of Consulting and Clinical Psychology*, 68, 898–908.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10, 349–367.
- Stout, R. L., Wirtz, P. W., Carbonari, J. P., & Del Boca, F. K. (1994). Ensuring balanced distribution of prognostic factors in treatment outcome research. *Journal of Studies on Alcohol*, 12, 70–75.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy*. Chicago: Aldine.
- Valle, S. K. (1981). Interpersonal functioning of alcoholism counselors and treatment outcome. *Journal of Studies on Alcohol*, 42, 783–790.
- VandeCreek, L., & Brace, K. (1991). Mandatory continuing education in the health professions. *The Journal of Training and Practice in Professional Psychology*, 5, 23–36.
- Woollard, J., Beilin, L., Lord, T., Puddey, I., MacAdam, D., & Rouse, I. (1995). A controlled trial of nurse counseling on lifestyle change for hypertensives treated in general practice: Preliminary results. *Clinical and Experimental Pharmacology and Physiology*, 22, 466–468.
- Ziskin, J. (1970). *Coping with psychiatric and psychological testimony*. Beverly Hills, CA: Law and Psychology Press.

Received August 1, 2003

Revision received December 18, 2003

Accepted February 20, 2004 ■