

What is Seen Through the Looking Glass: The Impact of Training on Practitioner Self-Rating of Motivational Interviewing Skills

Bryan Hartzler, John S. Baer, Chris Dunn, Dave B. Rosengren and Elizabeth Wells

University of Washington, Seattle, USA

Abstract. Training efforts for evidenced based treatments require evaluation, yet the value of practitioner self-reports of skills acquisition has been questioned. Thus, a key issue concerns how accurately practitioners assess their own clinical skills. In the current study, 23 community practitioners participated in training of Motivational Interviewing (MI), completed standardized patient (SP) interviews before and after training, and provided self-ratings of MI elements after each interview. Interview recordings were later coded independently. Results suggest training contributed to: 1) reasonable agreement between practitioner and independent ratings; and 2) more effective use of MI, despite a tendency for practitioners to underestimate training gains. This micro-analysis of training documents initial skill gains along with increased practitioner self-awareness. Further, it exemplifies how practitioner self-ratings and objective skill assessment methods may be used in tandem to more fully describe practitioner learning.

Keywords: Motivational interviewing, practitioner skill assessment.

Introduction

Motivational interviewing (MI; Miller and Rollnick, 1991, 2002) is a client-centered yet directive counseling style that aims to assist clients in recognizing and resolving ambivalence about behavior change. This therapeutic approach rests on practitioner adherence to the following principles: 1) expression of empathy; 2) development of discrepancy between client values and behavior; 3) defusing of, or “rolling with”, client resistance; and 4) support of client self-efficacy for behavior change. This approach has been widely applied to a variety of clinical issues and garnered strong empirical support (for reviews, see Dunn, Deroo and Rivara, 2001; Hettema, Steele and Miller, 2005; Vasilaki, Hosier and Cox, 2006). Consequently, MI training is continually sought internationally by community service organizations as a cost-effective means of improving client outcomes.

Despite rapid diffusion of MI, a small but growing number of formal training evaluations suggest variable skill gains wherein many practitioners do not achieve desired competency levels (Baer et al., 2004; Catley et al., 2006; Miller & Mount, 2001; Miller, Yahne, Moyers, Martinez and Pirritano, 2004; Schoener, Madeja, Henderson, Ondersma and Janisse, 2006; Shafer, Rhode and Chong, 2004). It bears mentioning that variation in training response is not

Reprint requests to Bryan Hartzler, Alcohol & Drug Abuse Institute, University of Washington, Box 354805, 1107 NE 45th Street, Suite 120, Seattle, WA 98105–4631, USA. E-mail: hartzb@u.washington.edu

© 2007 British Association for Behavioural and Cognitive Psychotherapies

specific to the MI approach (Walters, Matson, Baer and Ziedonis, 2005). Still, these findings, and the continued, international interest in MI training, make this a salient domain in which to explore *how* and *how effectively* practitioners seek, learn, and implement evidence-based practices.

To do so, one must first consider how best to assess the impact of participation in formal training processes. Existing MI training literature has minimized the value of practitioner reports, citing discordance with 3rd party skill ratings (Miller et al., 2004; Miller and Mount, 2001). Indeed, practitioner self-reports of knowledge, skill, adoption, or intent to adopt may hold limited predictive validity. Miller and Mount's (2001) evaluation of an MI workshop is one example of such discrepant findings. The study documented increased trainee perception of MI proficiency (e.g. Likert ratings of statements like "I feel proficient and able to use MI in my practice", or "I am already using MI in my work") despite lack of corresponding change in practice behavior. On this basis, practitioner reports were characterized as overly optimistic, with additional caution levied that such optimism may serve as a disincentive for future learning. Yet, trainees' presumed lack of accurate self-perception appears in conflict with a fundamental assumption underlying MI training – that learning involves ongoing processes of self-evaluation. If trainees are presumed to be unaware of their skills or deficiencies, how are skills to develop?

Assessment methods that rely on practitioner ratings may understandably raise concern, given potential for response bias (e.g. social desirability, acquiescence). However, two recent investigations (Peterson, Baer, Wells, Ginzler and Garrett, 2006; Strang and McCambridge, 2004) offer methods that may counter such influences, and thereby maximize rating accuracy. Each study explored the predictive value of post-session practitioner ratings on the efficacy of an MI-based intervention, using ratings that were: 1) based on direct in-session experience; 2) elicited immediately after sessions; and 3) focused on specific MI-relevant concepts. In each instance, practitioner ratings predicted eventual clinical outcomes, though it is noteworthy that these ratings were oriented to client, rather than practitioner, behavior (e.g. global rating of client "engagement", number of client change statements). Of further note, neither study included a measure of practitioner confidence in their ratings of the frequency and/or quality of specific in-session therapeutic events. Whether practitioners can confidently and accurately assess their own behavior in the context of attempting to learn and implement an evidence-based practice (e.g. MI) remains an open question.

The described efficacy studies justifiably relied on practitioner ratings of sessions with actual clients. However, this methodological approach is laden with complications when applied to practitioner skill assessment in a context where treatment dissemination (rather than efficacy) is the aim. Though a direct and realistic test experience, the use of actual clinical sessions does not control for variation in client characteristics that may impact attempts by practitioners to implement a therapeutic approach, and thereby introduce unwanted measurement variance. An alternative is the use of standardized patients (SPs), typically laypersons with a background in acting who simulate practitioner-client encounters for teaching and testing purposes. Audio-recording of practitioner interviews with an SP have been increasingly utilized in MI training evaluations in recent years (Baer et al., 2004; Miller et al., 2004; Miller and Mount, 2001), and provide realistic and consistent stimuli for the measuring of clinical skills.

The current study was incorporated into the pilot phase of a dissemination trial testing methods of MI training at two substance abuse treatment facilities in the northwestern United States. This study involved the use of SP interviews, conducted both before and after training,

as a means of assessing the impact of the training process. Thus, current data include ratings of MI elements by practitioners and independent coders from these interviews. These parallel ratings provide a micro-analysis of what practitioners see through the looking glass when assessing their own skills, and how that image may change as a function of training. Study hypotheses focus on the impact of training on the extent to which practitioner self-evaluations become:

- 1) More consistent with those of independent reviewers;
- 2) More confidently derived;
- 3) Indicative of perceived (vs. actual) improvement in MI skill.

Method

Recruitment

The University of Washington Institutional Review Board (IRB) reviewed and approved all procedures. Participants were recruited sequentially from two substance abuse treatment facilities, and offered free MI training at their respective facility as well as corresponding credit units for continuing education. Both agencies were compensated at a flat rate (determined a priori) to help support staff participation during regular clinic hours. Interested staff provided informed consent, and agreed to complete training outcome assessments during the week prior to and the week following training in exchange for personal financial compensation (\$30 per assessment). It should be noted that staff persons at one agency were prohibited by local county regulations from accepting such remuneration for activities occurring during regular work hours.

Participants

Thirty practitioners initially agreed to participate ($N = 14$ at agency #1, $N = 16$ at agency #2), and 23 completed training and pre/post SP interviews ($N = 12$ at agency #1, $N = 11$ at agency #2). Reasons for attrition included extended absence due to personal/family illness, extended absence due to personal/family vacation, prospective retirement, competing personal and agency time demands, and diminished interest in the project. Demographic and background characteristics of the final sample are listed separately for each agency in Table 1. Given the relative similarity of the two sets of agency personnel on these background indices, the analyses reported herein were conducted across the full sample rather than separately for each agency.

Training process

Training processes were consistent at each treatment agency. In both settings, five 3-hour group sessions were facilitated by a pair of licensed psychologists, both members of the Motivational Interviewing Network of Trainers (MINT) with considerable trainer experience. Group sessions occurred at 2-week intervals, and included traditional training components (e.g. didactic presentations, group exercises, role-plays). Between sessions, participants also completed an individual skills-practice interview that was audio-recorded and reviewed by a

Table 1. Practitioner demographic and background characteristics by agency

Characteristic	Agency #1 (N = 12)	Agency #2 (N = 11)
Age ^a		
M:	47.8 (10.8)	50.0 (9.7)
Range:	30–61	34–68
Clinic experience ^a		
M:	11.5 (7.4)	12.3 (8.5)
Range:	2–24	2–26
Gender		
% Female:	92	73
% Male:	8	27
Ethnicity		
% Caucasian:	92	73
% Other (specified)	0	27 ^b
% Other (unspecified)	8	0
Professional orientation		
% CDP ^c	58	55
% Other counseling background	17	27
% Medical	8	9
% Administrative	17	9
Education ^d		
% High School diploma/GED	17	18
% Associate's degree	17	46
% Bachelor's degree	58	9
% Master's degree	8	27
Prior exposure to MI ^e		
% Yes	58	55
% No	42	45

Notes: ^avalues reflective of years;

^btwo practitioners self-identified as “Asian-American”, one self-identified as “Alaska Native/American Indian”;

^creflects state credentialing as a chemical dependency professional;

^dreflects highest degree attained;

^ereflective of having: a) attended an MI presentation, brief training, or 2-day workshop, b) read the Miller and Rollnick (1991, 2002) text or journal articles about MI, and/or c) watched MI training tapes.

trainer who provided brief written feedback. Thus, time was allotted within group sessions to engage trainees in discussion of interview experiences and processing of trainer feedback.

Measures

Training outcome assessments included a 20-minute interview with an SP who portrayed a recently referred client encompassing common background characteristics for agency clientele. Interviews were recorded, after which practitioners completed brief rating forms of global indices and behavior counts from the Motivational Interviewing Treatment Integrity scale (MITI; Moyers, Martin, Manuel, Hendrickson and Miller, 2005) and domains of client change

language adapted from the Motivational Interviewing Skills Code (MISC; Moyers, Martin, Catley, Harris and Ahluwalia, 2003). Additionally, interview recordings were coded by one of two independent reviewers who had received extensive training and supervision in recognizing and rating MI elements in therapeutic dialogue. Independent ratings encompassed the same sets of therapist and client dimensions (adapted from MITI and MISC) as rated by practitioners. To minimize impact of preconceived hypotheses about training effects, independent reviewers were “blinded” to the time point of interviews as well as to the practitioner and agency. Procedures were also employed to continually evaluate reviewers’ inter- and intra-rater reliability, in addition to the formal reliability analyses (detailed later) undertaken to determine appropriate indices for subsequent tests of study hypotheses.

Interview ratings relied on a consistent system of indices, whether rated by practitioner (retrospectively after the interview via a ratings form) or an independent reviewer (in real time during a subsequent review of interview recordings). This rating system included global indices and counts of therapist behavior, as well as counts of client change language.

Global ratings. Global ratings covered two broad therapist dimensions, each rated on a 7-point Likert scale (1 = Low, 7 = High). These were: 1) empathy, or the extent to which understanding of client perspective was conveyed; and 2) MI spirit, or the overall competence exhibited in hallmark characteristics of MI such as collaboration with the client, evocation of client ideas, and the supporting of client autonomy. Consistent with suggested use of the MITI system (Moyers et al., 2005), practitioners and independent reviewers derived ratings consistent with an overall, “gestalt” impression of how well or poorly the intent of the scale was met.

Therapist behavior counts. The frequency of six therapist behaviors was tallied. These were: 1) closed questions (CQ), or those eliciting a “yes”, “no”, or response within a restricted range (e.g. a number, a day of the week); 2) open questions (OQ), or those promoting client narrative and allowing a wider range of possible answers; 3) simple reflections (SR), or statements conveying understanding of client perspective with little added meaning or emphasis; 4) complex reflections (CR), or statements conveying understanding of client perspective with substantive added meaning or emphasis; 5) MI adherent behaviors (MIA), or behaviors thought to be particularly exemplary of MI (e.g. asking permission to advise, affirming client choice or self-direction); and 6) MI non-adherent (MINA), or behaviors thought to be particularly contrary to the MI approach (e.g. advising without permission, confrontation). Practitioners also endorsed confidence ratings using a 5-point Likert scale (1 = Not at all confident, 5 = Very confident) for their estimate of the respective frequency of each of these therapist behaviors.

Client change language counts. The frequency of six domains of client change language was also tallied. These were: 1) positive change talk (CT+), encompassing stated or implied desire, ability, reason, or need for change; 2) negative change talk (CT−), encompassing stated or implied desire, ability, reason, or need for the status quo; 3) neutral change talk (CT0), reflecting discussion of change with unclear valence; 4) positive commitment talk (C+), indicative of clear, stated intent to change; 5) negative commitment talk (C−), indicative of clear, stated intent to maintain status quo; 6) neutral commitment talk (C0), reflecting commitment discussion with unclear valence. Practitioners similarly endorsed confidence

ratings for frequency estimates in each domain using the same 5-point Likert scale described above for counted therapist behaviors.

Results

As a preliminary consideration, reliability of the independent reviewers was assessed in a subset of 15 interviews. Cichetti (1994) offers a rationale for the use of intra-class correlations (ICCs) as a more stringent psychometric means of evaluating reliability between observations over Pearson product-moment correlations (for which covariance of observations is more simply divided by the product of their standard deviations), namely that ICCs: 1) distinguish sets of observations based on both their ordering and level of agreement; and 2) correct for a level of test-retest agreement that might be expected by chance alone. Further, Cichetti (1994) proposes particular guidelines for evaluating ICCs when the observations pertain to assessment of clinical phenomena. ICCs for the two global ratings, Empathy and MI Spirit, were .46 and .52, respectively. ICCs for counted therapist behaviors were: CQ, .92; OQ, .93; SR, .62; CR, .44; MIA, .60; and MINA, .65. As outlined by Cichetti (1994), all values were within acceptable ranges (.40–.59 = fair, .60–.74 = good, .75–1.00 = excellent) and indices were retained for subsequent analyses. ICCs for client change language were: CT+, .41; CT–, .68; CT0, .14; C+, .18; C–, .00; and C0, .00. Closer inspection of the data suggested poor reliability for CT0 was due to persistent reviewer disagreement, whereas commitment talk categories (e.g. C+, C–, C0) suffered from low base rate of occurrence. Thus, only CT+ and CT– indices were retained. Intra-rater reliability was also assessed through re-coding of 10% of each reviewer's work. Reviewer-specific ICCs were computed for the 10 retained indices (e.g. Empathy, MI Spirit, CQ, OQ, SR, CR, MIA, MINA, CT+, and CT–), with acceptable values across indices (reviewer A, .70–1.00; reviewer B, .72–.97) suggestive of consistent interview rating over time.

Subsequent analyses addressed each of the three aforementioned study hypotheses. The first hypothesis, focused on consistency between the ratings of practitioners and independent reviewers, was addressed by way of bivariate correlations at the level of individual indices. Additionally, Fisher *R to Z* transformations were applied to address whether pre-to-post training changes in the strength of these correlations achieved statistical significance. The second hypothesis, focused on practitioners' confidence in their own ratings, relied on a series of repeated-measures multivariate analysis of variance (MANOVA) to detect pre- to post-training change in reported confidence for the estimated frequency of therapist behaviors and domains of client change language. The third hypothesis, focused on pre-to-post training change in perceived vs. actual MI skill, also relied on a pair of repeated-measures MANOVA for each of two key summary scores of MI skill, computed in parallel from the ratings of independent reviewers (e.g. actual MI skill) and those of the practitioners (e.g. perceived MI skill).

Consistency between ratings of independent reviewers and practitioners

Bivariate correlations were computed from pre- and post-training SP interviews between the ratings of the independent reviewers and those of the practitioners for global therapist indices, counted therapist behaviors, and counted domains of client change language (see Table 2). Correlations at pre-training were variable ($r = -.03-.43$), and did not reach statistical significance in most cases. As a point of reference, Cohen (1988) offers the following

Table 2. Consistency between independent and practitioner ratings

Index	Pre-training	Post-training	Z-difference score
Therapist global markers			
Empathy:	.016	.289	1.29
MI Spirit:	−.032	.464*	2.15*
Therapist behavior counts			
CQ:	.025	.717***	3.63***
OQ:	.425*	.516*	1.05
SR:	.221	.458*	1.24
CR:	.195	.484*	1.50
MIA:	.258	.543**	1.57
MINA:	.322	.501*	1.03
Client language counts			
CT+:	.433*	.568**	.568**
CT−:	.015	.572**	2.88**

Notes: Indices reflect as follows: CQ = practitioner closed questions, OQ = practitioner open questions, SR = practitioner simple reflections, CR = practitioner complex reflections, MIA = practitioner MI-adherent behaviors, MINA = practitioner MI non-adherent behaviors, CT+ = client pro-change statements, CT− = client anti-change statements.

Values in “Pre-Training” and “Post-Training” columns reflect bivariate correlations for that index between ratings of practitioner and an independent reviewer.

Correlations based on sample of 23 practitioners who completed MI training process and both SP interviews.

Z-difference scores reflect temporal change in magnitude of correlations, as assessed via *R* to *Z* transformations ****p* < .001, ***p* < .01, **p* < .05.

classification to determine the strength of a Pearson product-moment correlation (*r*): .10–.29, *small*; .30–.49, *medium*; .50–1.00, *large*. Per Cohen’s criteria, practitioner-reviewer ratings for four indices (global empathy, global MI spirit, CQ and CT−) were uncorrelated, whereas pre-training correlations for other MI elements were small (SR, CR, and MIA) to medium (OQ, MINA and CT+). Taken together, this suggests modest and inconsistent agreement between the ratings of practitioners and independent reviewers prior to training.

Post-training correlations between ratings evidenced a different pattern, with larger correlation coefficients observed across all MI elements and statistically significant positive association for nine of ten indices (see Table 2). According to Cohen’s (1988) criteria, post-training correlations between practitioner-reviewer ratings were large for six indices (CQ, OQ, MIA, MINA, CT+ and CT−), medium for three indices (global MI spirit, SR and CR), and small for just one index (global empathy).

To formally test the impact of training on associations between practitioner-reviewer ratings, Fisher *R* to *Z* transformations were applied. This procedure, described in detail elsewhere (Fisher, 1970), formally quantifies the difference between the magnitude of two correlations by transformation of each to a standardized metric (e.g. *Z* score) from which a difference is calculated and interpreted in relation to sample size and what might be expected by chance alone. Fisher *R* to *Z* transformations were applied to each of the 10 noted MI elements, with the magnitude of correlation for practitioner-reviewer ratings prior to training compared with that observed following training. Applying a 95% confidence interval as a criterion (for which

a Z -score difference of 1.96 is required for statistical significance), three elements met this standard: 1) MI Spirit, $Z=2.15$, $p < .05$; 2) CQ, $Z=3.63$, $p < .05$; and 3) CT-, $Z=2.88$, $p < .05$. The remaining seven elements did not meet this criterion, though all were positive and 6 of the 7 achieved greater than a 1.0 Z -score difference (see Table 2).

Confidence estimates for practitioner ratings

A series of repeated-measures MANOVA analyses were next employed to assess change in practitioner confidence in their frequency estimates for the six counted therapist behaviors (e.g. CQ, OQ, SR, CR, MIA and MINA). The series of MANOVA revealed minimal change in practitioner confidence for frequency estimates of these behaviors (all F -values < 1.22 , all p -values $> .27$). Similarly, repeated-measures MANOVA was employed for confidence ratings of practitioners' frequency estimates for each of the two domains of client change language (e.g. CT+, CT-). MANOVA again indicated minimal change in practitioner confidence as a function of training (both F -values $< .20$, both p -values $> .65$). Closer inspection of the means for practitioner confidence ratings revealed a consistent pattern wherein moderate confidence was endorsed for frequency estimates across all indices both prior to and after training.

Change in actual and perceived MI skill as a function of training

Repeated-measures MANOVA next assessed temporal change in: 1) independent ratings, suggestive of actual skill in performing MI; and 2) practitioner ratings, suggestive of perceived MI skill. One's ratio of reflections to questions (e.g. R:Q) has been offered as a meaningful conceptual index of MI skill (Moyers et al., 2003; Baer et al., 2004). Similarly, the percentage of one's behaviors thought to be emblematic of the MI approach (e.g. % MI-Adherent behavior) has also been utilized to reflect MI skill (Moyers et al., 2006). Consequently, R:Q and % MI-Adherent behavior were chosen as proxies for assessing actual and perceived MI skillfulness, and were computed separately from the independent and practitioner ratings.

MANOVA detected a marked increase in actual R:Q, $F(1,22) = 33.95$, $p < .001$. As illustrated in Figure 1, the mean R:Q just after training ($M = 1.07$, $SD = .64$) more than doubled that observed before training ($M = .49$, $SD = .33$). Closer inspection of the data revealed that 20 of the 23 practitioners improved R:Q by a standard deviation (SD) or more, and that 10 of these improved by at least two SD s. MANOVA also detected an increase in actual % MI-Adherent behavior, $F(1,22) = 8.16$, $p < .01$; Figure 2 illustrates that the mean % MI-Adherent behavior at post-training ($M = .70$, $SD = .44$) doubled that observed before training ($M = .35$, $SD = .45$). Inspection of the data indicated that 11 practitioners improved by an SD or more, and that seven improved by two SD s. Of note, analyses of R:Q and % MI-Adherent behavior that included practitioner background variables (e.g. age, clinic experience, gender, professional orientation, education, prior MI exposure) as covariates did not impact the described outcomes (all F -values < 2.01 , p -values $> .17$). The collective pattern suggests training resulted in marked improvement in actual MI skill.

In contrast, MANOVA revealed nonsignificant increases in perceived R:Q and % MI-Adherent behavior. As noted for prior analyses, both MANOVAs were re-computed with practitioner background variables included as covariates, and with no apparent impact (all F -values < 1.43 , all p -values $> .24$). As Figure 1 illustrates, mean perceived R:Q just after training ($M = .76$, $SD = .48$) only marginally improved relative to that perceived before training

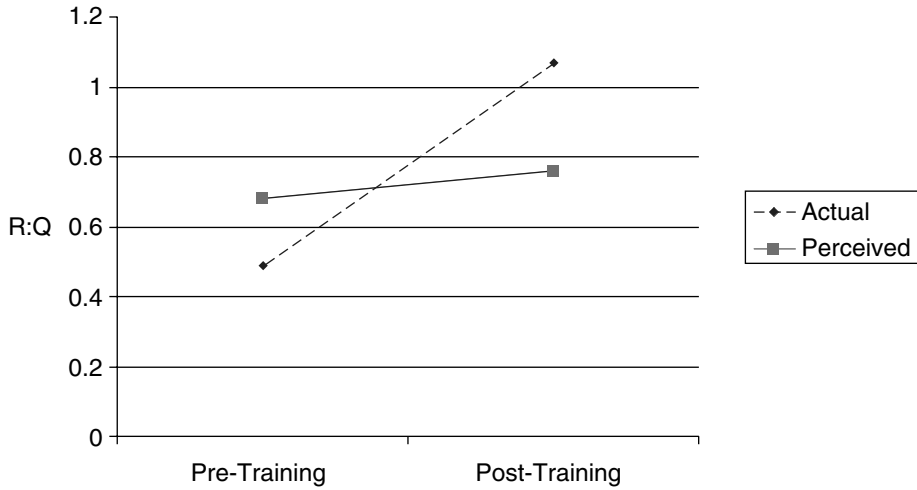


Figure 1. Impact of training on actual and perceived ratio of reflections to questions
Notes: R:Q reflects practitioner ratio of reflections to questions (Total Reflections/Total Questions) as a proxy of MI skill.
 Actual R:Q was derived from independent ratings of SP interviews.
 Perceived R:Q was derived from practitioner ratings of SP interviews.

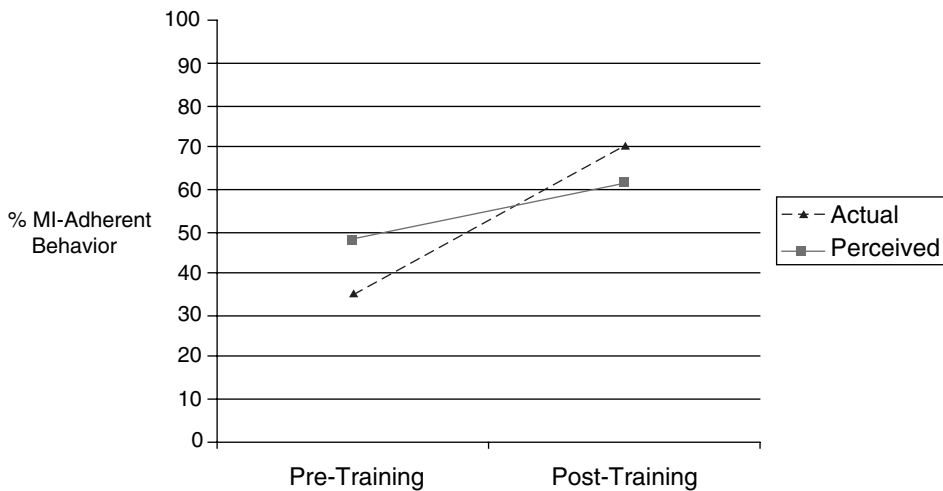


Figure 2. Impact of training on actual and perceived percentage of MI-adherent behaviors
Notes: % MI-Adherent behavior is a proxy of MI skill computed as the proportion of practitioners' behavior that received an MIA code from among all behaviors receiving either an MIA or MINA code (e.g. MIA/(MIA + MINA)).
 Actual % MI-Adherent behavior was derived from independent ratings of SP interviews.
 Perceived % MI-Adherent behavior was derived from practitioner ratings of SP interviews.

($M = .68$, $SD = .31$). Closer inspection of the data indicated an increase of a SD or more by 7 of 23 practitioners, and an increase of at least two SD s in only four cases. Figure 2 illustrates that perceived % MI-Adherent behavior just after training ($M = .62$, $SD = .35$) marginally improved over that perceived before training ($M = .48$, $SD = .27$). Data revealed an increase of a SD or more by 9 of 23 practitioners, and an increase of at least two SD s in three cases. The collective pattern suggests that, on average, practitioners overestimated skills before training but did not fully recognize initial skill gains when asked to rate these after training had concluded.

The pattern of findings concerning perceived MI skill was somewhat unexpected, and consequently was further scrutinized. The R:Q was chosen as the focus, given that actual skill gains were most robustly and reliably observed in this index. As noted, 20 practitioners evidenced a marked increase in actual R:Q (of a SD or more) as a function of training. Of this group, just seven perceived R:Q improvement of similar magnitude. The remainder was comprised of two practitioners who perceived R:Q improvement of lesser magnitude, nine who perceived no appreciable change in R:Q, and two who perceived meaningful *deterioration* in R:Q as a function of training. Of further note, the three remaining practitioners in this sample (e.g. those not evidencing a meaningful increase in actual R:Q) also perceived a deterioration in R:Q. Thus, the current data, although based on a small sample and an uncontrolled study design, provide little support for the notion that practitioners are prone to overly optimistic impressions of their skills after training. Instead, practitioners in the current sample, on average, in most cases, and across levels of performance, were prone to underestimate their actual skill gains.

Discussion

This study explored the immediate impact of training on self-ratings by community practitioners concerning their MI skill. This micro-analysis of practitioners' subjective learning experience relied on post-session practitioner ratings of MI elements, which were compared with parallel ratings derived by independent reviewers using indices from established MI coding systems (Moyers et al., 2003, 2005). Practitioner rating forms elicited ratings that were experientially-based, immediately-derived, and behaviorally-specific. Study results bear on two goals of technology transfer efforts, namely the extent to which training influences: 1) the accuracy and confidence of practitioner self-perceptions about *what they do*; and 2) perceived and actual practitioner skill during attempts to perform *what they learn*.

Accuracy and confidence in practitioner self-ratings

Unsurprisingly, associations of parallel ratings by practitioners and independent reviews were variable prior to training, exhibiting lack of association for several indices. This may be unremarkable, given that practitioners had not yet engaged in the training process. A different pattern of associations emerged after training, with medium-to-large correlations observed between the ratings of practitioners and independent reviewers across nine of ten MI elements. Of note, the Fisher Z transformation procedure revealed statistically significant increase in the congruence of practitioner-reviewer ratings of MI spirit, number of closed questions, and number of negative client change statements as a function of training. If one accepts that independent reviewers provide valuable insights about treatment fidelity (Moyers et al., 2005),

then the ability of practitioners to reach like-minded conclusions about their work – that is, to more accurately self-evaluate – is of value, irrespective of the indicated skill level. Perhaps this impact of a learning process may be considered, in and of itself, a marker of initial training success.

Congruence in the ratings of practitioners and reviewers is more impressive when one considers the diversity of MI elements that were rated. Among these were: 1) specific practitioner techniques (e.g. offering reflective listening statements) historically emphasized in the context of learning this therapeutic approach (Rollnick and Miller, 1995); 2) specific client behaviors (e.g. stating a desire for change) suggested to be proxies for clinical outcomes (Amrhein, Miller, Yahne, Palmer and Fulcher, 2003); and 3) broad practitioner interpersonal dimensions (e.g. the capacity for empathic understanding) that are increasingly recognized as predictors of successful implementation of MI (Moyers, Miller and Hendrickson, 2005). Clearly, there is great potential for a breadth of experiential learning via structured training processes.

One area in which practitioners appeared to benefit little was in confidence in the ratings they provided. Despite improvement in the accuracy of practitioner ratings, the strength of the accompanying confidence estimates changed little. In fact, practitioners uniformly endorsed moderate confidence across MI elements and over time. Perhaps confidence in the accuracy of self-ratings may be more variable and sensitive to change when applied to broader, MI-relevant interpersonal qualities (e.g. empathy), or to the absence/presence of individual MI elements (e.g. affirmations) rather than retrospective recall of their actual frequency per se. It is also conceivable that practitioner confidence about the accuracy of self-evaluations of one's skills and performances is an area wherein current training efforts can improve.

Actual and perceived MI skill

Two indices of MI skill – R:Q and % MI-Adherent behavior – were computed separately from the ratings of independent reviewers and of practitioners. Analyses revealed large initial training effects wherein actual MI skill improved, both substantially for the group overall and with reasonable reliability across individual practitioners. Interestingly, these training effects were not predicted by any of a variety of practitioner background variables (e.g. age, gender, experience, education, prior exposure to MI). The magnitude of these initial training effects approached that recently reported for MI training methods that similarly included structured practice and feedback (Miller et al., 2004), and is consistent with training literature for other therapeutic approaches utilized by this population of practitioners (Sholomskas et al., 2005).

Increases in perceived MI skill were subtle, failing to reach statistical significance and with only a minority of the sample perceiving enhanced skills. This was surprising, given the magnitude of the initial training effects observed in practitioners' actual skills, though these results were similarly uninfluenced by any of the practitioner background variables previously noted. Perhaps the skills-practice and feedback components of this training format offered more explicit and personalized challenges to practitioners, and thereby diminished perceived skill gains by reinforcing a need for continued practice. Alternatively, the pattern of effects may capture an important distinction between the skill set required to perform MI versus that required to codify it. For example, the behavioral task of responding to a client by listening reflectively in-session may be more readily achieved by some practitioners than the particularly linguistic, perhaps meta-cognitive task of counting and categorizing units of speech intended to represent aspects of reflective listening. Codifying therapeutic dialogue is

complex, and lends itself to interpretive quandaries about subjective meanings and implicit intents of either involved party. Whether such quandaries occur in the context of education, evaluation, or clinical service, they present ongoing challenges to our collective understanding of this therapeutic approach.

Caveats and limitations

For a number of reasons, caution is warranted in interpreting study results. Prominent among these is a small sample size ($N=23$), which limits confidence about generalizability of the observed effects. It is noteworthy that the sample consisted of similarly diverse staffs from two community-based agencies, whose professional titles and typical workplace duties encompassed a range of counseling, medical, and/or administrative services in support of providing substance abuse treatment. Further, assessment and training processes were conducted sequentially at the two facilities over approximately 3 months per agency. Still, there is limited assurance that this sample is representative of substance abuse treatment practitioners at large, let alone those in other helping professions and/or that work with different clinical issues or populations. The current sample size similarly limited the statistical power available to *observe* effects. It is quite possible that this obscured detection of change in some domains, given the stringency of certain described data analytic methods (e.g. Fisher R to Z transformations). Clearly, generalizability of study findings would be informed by attempts at replication with a larger sample of practitioners.

Other significant caveats are the absence of both a control group and of a follow-up assessment, both of which would provide measures of a number of potential influences to better account for the temporal changes detected in perceived and actual MI skill. It is conceivable that forms of assessment reactivity (e.g. socially-desirable responding, practice effects), for example, played a role in the observed temporal changes in indices of actual and/or perceived MI skill, or that such reactivity might predict temporal change in practitioner confidence in their self-ratings (though this was not observed). It should be noted that the temporal changes that were observed typically coincided with the intent and content of training; however, the absence of an untrained control group does limit the certainty of the conclusions one might draw about the impact of these training processes on practitioners' demonstrated MI skills and their perceptions of these skills. It is also conceivable that any of the described impacts of training may deteriorate over time, as indicated in other dissemination trials (Baer et al., 2004, Miller and Mount, 2001).

An additional caveat concerns the use of SP interviews as a means of evaluating training processes. Practicalities inherent in this study design prevented the SP from remaining "blind" to whether interviews occurred before or after training; thus, any biases or expectations the SP may have held about training effects may have influenced her conduct, and therefore the therapeutic process, during interviews. Further, SP characters may have lacked ecological validity in some respects and consistency of presentation in others. To minimize such threats to validity, study personnel: 1) collaborated with agency directors to design characters reflective of their clientele; 2) hired an SP who was experienced in the conduct of university-based work with healthcare providers; and 3) provided extensive training to the SP to maximize authenticity and consistency in the portrayal of her "characters". Even so, there remains the possibility that the noted concerns unduly influenced aspects of the described SP interviews from which the ratings were derived.

Implications and conclusions

Limitations notwithstanding, the current study offers insights into the learning processes undertaken by community practitioners as they engage in training processes to learn an evidence-based practice. Though admittedly small and uncontrolled, the study provides a micro-analysis of the initial impact of training processes on community practitioners' skills, their perceptions of those skills, and their confidence in those perceptions. Moreover, it describes a simple method of assessing MI skill through a brief written rating form that prompts practitioners to self-evaluate performances in a variety of defined MI domains following a therapeutic interaction.

In the spirit of this client-centered therapeutic approach, the described self-rating system may reflect a move toward a more "trainee-centered" form of practitioner assessment. The recent emergence of MI coding systems (Madson, Campbell, Barrett, Brondino and Melchert, 2005; Moyers et al., 2003, 2005) seems to be, at least in part, in response to the notion that practitioners cannot reliably report what they know or don't know, what they learn or don't learn, or what they do or don't do. It has been argued that increased practitioner confidence may be a detrimental initial training outcome (Miller and Mount, 2001); that is, if one's self-efficacy to perform MI is undeservedly high, then one may be less prone to seek further training needed to improve what may be under-developed skills. Though a reasonable and practical concern, few, if any, practitioners in the current sample endorsed inflated perceptions of the skills they exhibited during SP interviews occurring the week following training. On the contrary, the current data suggest that, when assessed in an experiential and behaviorally-specific manner, self-ratings of MI skills may be less apt to change as a function of training than ratings of those same skills by independent reviewers. In fact, the current data raise the question of whether there is a need to approach MI training in a manner that more effectively builds practitioner self-awareness and self-efficacy for its implementation.

Though Miller and Mount (2001) have questioned the veracity of practitioner self-reports, perhaps this is not an issue of *whether* practitioners can evaluate their own skills, but rather *how* we ask them about their perceived skills, abilities, or performances. Those in research and training communities may risk throwing the baby out with the bathwater if we discount what practitioners can tell us about the impact that training programs and experiences may have on their therapeutic skills and practices. Instead, it seems a primary challenge is to design and provide practitioners with learning processes that strike a balance between fostering self-awareness and supporting self-efficacy, much in the spirit of the therapeutic process that practitioners are asked to undertake with their clients when implementing MI (Rollnick and Miller, 1995). The current study provides one small example of how "objective" and "subjective" forms of practitioner skill assessment may be utilized in tandem. And in so doing, perhaps it provides a fuller picture of how the image in the looking glass changes as practitioners consider and explore, understand and appreciate, practice and self-evaluate, and adopt and adapt the spirit and techniques of this therapeutic approach.

Acknowledgements

This research was supported by NIDA 1 R01 DA016360-01A1, *Development and Evaluation of Context Tailored Training* (John Baer, PI). The authors wish to thank Jeretta Scott for her expertise in portraying substance abuse treatment clientele during the described standardized

patient interviews, Andrew Slade and David G. Peterson for their efforts in assisting with data collection and coding activities, and the participating agencies and practitioners for their efforts.

References

- Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M. and 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., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L. and Hartzler, B.** (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence, 73*, 99–106.
- Catley, D., Harris, K. J., Mayo, M. S., Hall, S., Okuyemi, K. S., Boardman, T. and Ahluwalia, J. S.** (2006). Adherence to principles of motivational interviewing and client within-session behavior. *Behavioural and Cognitive Psychotherapy, 34*, 43–56.
- Cicchetti, D. V.** (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized instruments in psychology. *Psychological Assessment, 6*, 284–290.
- Cohen, J.** (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Dunn, C., Deroo, L. and Rivara, F. P.** (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. *Addiction, 96*, 1725–1742.
- Fisher, R. A.** (1970). *Statistical Methods for Research Workers* (14th ed.). Edinburgh: Oliver and Boyd.
- Hettema, J., Steele, J. and Miller, W. R.** (2005). Motivational Interviewing. *Annual Review of Clinical Psychology, 1*, 91–111.
- Madson, M. B., Campbell, T. C., Barrett, D. E., Brondino, M. J. and Melchert, T. P.** (2005). Development of the motivational interviewing supervision and training scale. *Psychology of Addictive Behaviors, 19*, 303–310.
- Miller, W. R. and 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. and Rollnick, S.** (1991). *Motivational Interviewing: preparing people to change addictive behavior*. New York: Guilford Press.
- Miller, W. R. and Rollnick, S.** (2002). *Motivational Interviewing: preparing people for change* (2nd ed.). New York: Guilford Press.
- Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J. and Pirritano, M.** (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology, 72*, 1050–1062.
- Moyers, T. B., Manuel, J. K., Wilson, P., Hendrickson, S. M. L., Talcott, W., Durand, P. and Cedro, K. M.** (2006). *Training Behavioral Health Providers in Motivational Interviewing: initial results from the TEAM study*. Poster presented at the 11th International Conference on Treatment of Addictive Behaviors, Santa Fe, New Mexico, February.
- Moyers, T. B., Martin, T., Catley, D., Harris, K. J. and 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.
- Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L. and Miller, W. R.** (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment, 28*, 19–26.
- Moyers, T. B., Miller, W. R. and Hendrickson, S. M. L.** (2005). How does motivational interviewing work? Therapist interpersonal skill as a predictor of client behavior with motivational interviewing sessions. *Journal of Consulting and Clinical Psychology, 73*, 590–598.

- Peterson, P. L., Baer, J. S., Wells, E. A., Ginzler, J. A. and Garrett, S. B.** (2006). Short-term effects of a brief motivational intervention to reduce alcohol and drug risk among homeless adolescents. *Psychology of Addictive Behaviors*, 20, 254–264.
- Rollnick, S. and Miller, W. R.** (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, 23, 325–334.
- Schoener, E. P., Madeja, C. L., Henderson, M. J., Ondersma, S. J. and Janisse, J. J.** (2006). Effects of motivational interviewing training on mental health therapist behavior. *Drug and Alcohol Dependence*, 82, 269–275.
- Shafer, M. S., Rhode, R. and Chong, J.** (2004). Using distance education to promote the transfer of motivational interviewing skills among behavioral health professionals. *Journal of Substance Abuse Treatment*, 26, 141–148.
- Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F. and Carroll, K. M.** (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, 73, 106–115.
- Strang, J. and McCambridge, J.** (2004). Can the practitioner correctly predict outcome in motivational interviewing? *Journal of Substance Abuse Treatment*, 27, 83–88.
- Vasilaki, E. I., Hosier, S. G. and Cox, W. M.** (2006). The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. *Alcohol & Alcoholism*, 41, 328–335.
- Walters, S. T., Matson, S. A., Baer, J. S. and Ziedonis, D. M.** (2005). Effectiveness of workshop training for psychosocial addiction treatments: a systematic review. *Journal of Substance Abuse Treatment*, 29, 283–293.