4-1-2018

Assessment of the Relationship between a Written Measure of Empathy and an Independently Rated Interview of Motivational Interviewing

Jennifer L. Smith
Columbia University

P. Bertone
City University of New York

Kenneth M. Carpenter
Columbia University

R. Morgan Wain
Columbia University

Mei Chen Hu
Columbia University

Journal of Substance Abuse Treatment 87 (2018) 56–63

Contents lists available at ScienceDirect

Journal of Substance Abuse Treatment

Assessment of the relationship between a written measure of empathy and an independently rated interview of Motivational Interviewing

Smith J.L. [a,b,*], Bertone P. [c], Carpenter K.M. [a,b], Wain R.M. [a], Hu M.-C. [b], Amrhein P.C. [b,d], Nunes E.V. [a,b]

[a] NYS Psychiatric Institute, Division of Substance Use Disorders, NY, NY 10032, United States
[b] Columbia University, Department of Psychiatry, NY, NY 10021, United States
[c] Center for HIV Educational Studies & Training, Hunter College, CUNY, NY, NY 10018, United States
[d] Montclair State, Montclair, NJ 07043, United States

ARTICLE INFO

Article history:
Received 16 March 2017
Received in revised form 11 December 2017
Accepted 11 December 2017

ABSTRACT

Motivational Interviewing (MI) is an evidence-based practice shown to be effective when working with people in treatment for substance use disorders. However, MI is a complex treatment modality optimized by training with feedback. Feedback, assessment and monitoring of treatment fidelity require measurement, which is typically done using audiotaped sessions. The gold standard for such measurement of MI skill has been an audiotaped interview, scored by a rater with a detailed structured instrument such as the Motivational Interviewing Treatment Integrity 2.0 (MITI 2.0) Coding System (Moyers, et al., 2005). The Helpful Responses Questionnaire (HRQ) (Miller, Hedrick, & Orlofsky, 1991) is a pen-and-paper test of empathy (a foundational MI skill) that does not require an audiotaped session. A randomized trial of three different regimens for training counselors in MI (live supervision using Teleconferencing, Tape-based supervision and Workshop only) (Smith et al., 2012) offered the opportunity to evaluate the performance of the HRQ as a measure of MI ability, compared to the several MITI 2.0 global scores and subscales. Participants were counselors (N = 97) working at community-based substance use treatment programs, whose MI proficiency was measured at four time points: baseline (before an initial 2-day MI workshop), post-workshop, 8 weeks post-workshop (i.e., post-supervision), and 20 weeks post-workshop with both MITI 2.0 and HRQ.

HRQ total scores correlated significantly with the Reflection to Question Ratio from the MITI 2.0 at post-workshop (r = 0.33), week 8 (r = 0.34), and week 20 (r = 0.38), and with the Spirit (r = 0.32) and Empathy (r = 0.32) global scores at week 20. Correlations of HRQ with other MITI 2.0 subscales and time points after workshop were small and not significant. As predicted, HRQ scores differed between training conditions ($X^2(2) = 7.88$, p = 0.02), with counselors assigned to live supervision achieving better HRQ scores than those in Workshop only. In summary, HRQ is a modestly accurate measure, mainly of the Reflection to Question Ratio, considered a core marker of MI skill. It is sensitive to training effects and may help identify counselors needing more intensive supervision. Given its ease of administration and scoring, HRQ may be a useful marker of MI skill during training efforts.

1. Introduction

The increasing demand for evidence-based practices (EBPs) in psychotherapy has required focused translation of research findings into clinical practice (Baer et al., 2004; Madson & Campbell, 2006). Motivational Interviewing (MI), a popular EBP frequently used in substance abuse treatment, is defined as a client-centered, collaborative but still directive counseling style "paying particular attention to the language of change", aimed at bolstering the person's own motivation to change specific behaviors (Miller & Rollnick, 2013). However, MI comprises a complex set of skills that are not easy to master (Forsberg, Forsberg, Lindqvist, & Helgason, 2010) so training requires valid measurement of EBP fidelity and skill. Research has demonstrated that MI workshops followed by feedback and supervision result in superior MI skill over workshop training alone (deRoten, Zimmermann, Ortega, & Despland, 2013; Forsberg et al., 2010; Martino, Gallon, Ball, & Carroll, 2008; Miller & Rose, 2009; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; Smith et al., 2012). Supervision should ideally include the gold standard of feedback from objective assessment of audiotaped interviews, either with real patients, actors simulating patients, or computer-simulated patients (Rosengren, Hartzler, Baer, Wells, & Dunn, 2008). However, audiotaping can be difficult to implement in

* Corresponding author at: NYS Psychiatric Institute, 1051 Riverside Drive, Room 3732, Box 120, New York, NY 10032, United States.
E-mail address: jennifer.smith@nyspi.columbia.edu (J.L. Smith).

https://doi.org/10.1016/j.jsat.2017.12.004
0740-5472/

community practice settings (Barwick, Bennett, Johnson, McGowan, & Moore, 2012; Bennett, Moore, et al., 2007; Bennett, Roberts, et al., 2007; Madson & Campbell, 2006). Reasons for this may include time constraints, limited financial and staffing resources (Olmstead, Carroll, Canning-Ball, & Martino, 2011), restricted access to technology, and counselors' reluctance to submit audiotapes for review (Baer et al., 2004; Vachon et al., 1995). Another significant challenge to successful implementation of MI at community-based substance abuse treatment programs is the fact that counselors view themselves as fairly proficient after initial training but self-report is a poor predictor of actual in-session competence (Carroll, Nich, & Rounsaville, 1997; Decker & Martino, 2013; Martino, Ball, Nich, Frankforter, & Carroll, 2009; Miller et al., 2004; Miller & Mount, 2001; Wain et al., 2015; Walters, Matson, Baer, & Ziedonis, 2005). It follows that self-report cannot be considered an accurate source of MI skill assessment. In this context, the need for reliable, valid, and easily administered tools to evaluate counselor skill in MI is crucial to establishing appropriate training protocols.

A simpler, yet objective, means of assessing MI skill, like the Helpful Responses Questionnaire (HRQ) - a pen-and-paper assessment intended to measure empathy (one of the foundational MI skills), would be of significant benefit to clinical supervisors and investigators in their efforts to train and disseminate MI widely among community-based counselors and treatment systems. Counselors completing the HRQ are asked to provide brief narrative responses to six clinical scenarios, which may lend insight into the counselors' understanding of and proficiency in MI. The HRQ is easy to score, can be used repeatedly over time, and does not require that an interview be recorded. Prior uses of the HRQ suggest it is a valid and reliable measure (Hartzler, 2015 (note: HRQ-CM); Miller, Hedrick, & Orlofsky, 1991) and therefore one possible candidate. If this questionnaire were an accurate assessment of MI ability, community-based treatment programs could satisfy their pressing need for simple and objective measures of MI competence by using it on a regular basis for MI supervision.

The HRQ has been widely used in diverse training settings and disciplines, including several adaptations of the scale: the HRQ-CM, to assess core communication skills in the delivery of contingency management (Hartzler, 2015); the WRQ and the ORQ, modified to evaluate employees in the juvenile justice system and correction officers, respectively (Doran, Hohman, & Koutsenok, 2011). The HRQ has also been adapted to teach MI to primary care staff for promoting various healthcare behaviors (Cucciare et al., 2012), teaching engineers learning vocational training (Klonek & Kauffeld, 2015); and it has been adapted to assess MI skill acquisition with internal medicine residents (Childers et al., 2012), case managers and program managers (Shafer, Rode, & Chong, 2004), and in the training of medical students (Daeppen et al., 2012). Some researchers have also looked at the potential for using the HRQ as a screening tool to segregate clinicians who come into training with high empathy, from those who do not and could thus further benefit from training that targets empathy skills (Stein, Clair, Soenksen-Bassett, Martin, & Clarke, 2015). In total, this widespread application of the HRQ in community settings highlights the need for quick and easy ways to implement measures of MI skill acquisition.

Still, questions exist about whether the HRQ is equivalent to observer-rated fidelity measures. Other studies have looked at the concurrent validity of the HRQ with similar measures of MI skill acquisition. For instance, Decker, Nich, Carroll, and Martino (2014) examined correlations between the HRQ and the scale they developed to measure empathy, the Therapist Empathy Scale (TES). The TES and the HRQ scores were not significantly correlated at any assessment points (Decker et al., 2014), and because the TES did indeed have a strong significant association with the MI integrity rating scores of the Independent Tape Rater Scale (ITRS), the authors concluded that written tasks (such as the HRQ) are inadequate to measure counselor empathy. Conversely, Baer et al., 2012, did find significant correlations between the HRQ and a measure that also uses vignettes depicting problems with substance use, i.e., the Computer Assessment of Simulated Patient Interviews (CASPI) (Baer et al., 2012).

There are limitations on the HRQ as a measure of MI skill, compared to an interview-based instrument like the MITI. First, the HRQ is a written measure in which participants are asked to imagine and write what they would say in response to a hypothetical patient statement; thus, a person's response is limited by their writing skills. Decker et al. (2013) questioned the equivalency—and thus the interchangeability—of real versus role-played sessions: they found low correspondence between treatment integrity ratings of real versus role-played sessions—counselors do not seem to show the same ability to use MI in the different scenarios of real and role-played sessions. Thus, it is logical to question the interchangeability between written responses to a written vignette of a simulated patient, and in-session responses in a clinical situation. Additionally, the MITI measures several facets of mi competence while the HRQ is intended to measure only one component of MI. In sum, the nature of the relationship between HRQ and gold standard independent observer session assessment is still unclear; therefore, we tested the relationships between HRQ and one of such gold standard measures, the MITI 2.0.

This report is based on a secondary analysis of a parent study that examined effects of supervision on the acquisition of MI skills (Smith et al., 2012). In the parent study, counselor-participants attended a two-day MI training workshop and were then randomized to one of three supervision conditions. One group received no supervision (i.e., Workshop only) and two groups received five weekly supervision sessions with an expert supervisor. Audiotaped training interviews conducted by the counselor-participants, with actors portraying clients at the counselor-participants' treatment sites, were the focus of supervision. In "Tele-Conferencing Supervision" (TCS), the supervisor used Teleconferencing technology to listen in to the interview live and give feedback in real-time. The other supervision group was "Tape," in which the recordings of training interviews were reviewed and discussed by the supervisor and counselor-participant on another day. The study showed that supervision, particularly the live supervision condition (i.e., TCS), was superior to workshop alone in fostering MI skill. As the primary assessment of MI proficiency, counselor-participants were audiotaped interviewing real patients at four time points: baseline (i.e., before workshop), post-workshop, 8 weeks post-workshop (i.e., after supervision was received), and 20 weeks post-workshop. The Motivational MITI 2.0 was used to score these assessment sessions, and to give counselor-participants feedback on their training interviews. At each assessment time point, counselor-participants also completed HRQ.

The goal in this secondary analysis was to evaluate the performance of the HRQ as a measure of MI skill, by comparing it with the MITI 2.0. We examined correlations between HRQ and MITI 2.0 scores across the four assessment times. Furthermore, we tested if the HRQ would detect the same effects of training condition as the MITI 2.0 did in the parent study (Smith et al., 2012). Our hypotheses were: 1) A positive association will be demonstrated between the HRQ and MITI 2.0 scores at each assessment point; and 2) Counselors who receive supervision will score higher on the HRQ than those who do not get supervision, a similar pattern as detected with the MITI 2.0 in the main outcomes analysis performed in the parent study. Additionally, to engage in the search for an assessment that could predict which counselors will benefit from MI supervision, we ran a ROC curve analysis to assess the utility of the HRQ in predicting MI proficiency.

2. Method

The Institutional Review Board of the New York State Psychiatric Institute and the Institutional Review Boards governing each of the participating treatment clinics approved all study procedures. A more detailed account of the design and procedures is presented in Smith et al. (2012).

2.1. Participants

2.1.1. Counselor-participants

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

2.1.2. Client-participants

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

2.1.3. Supervisors

Five doctoral-level clinical psychologists served as MI supervisors. All supervisors participated in a standard 2-day MI training workshop with a member of MINT and then submitted audiotaped clinical interactions that were rated for proficiency by the MINT trainer. All supervisors had to score at the competency level or higher on all MITI indices (version 2.0; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005) before supervising the counselor-participants. Several supervisors traveled to CASA at the University of New Mexico for MITI 2.0 training. All supervisors received training in the use of the MITI 2.0 for supervision purposes and met routinely to discuss coding questions.

2.2. Training procedures

2.2.1. Recruitment and assessment

Study personnel presented the training project to the clinic treatment staff, and a member of the research team interviewed potential counselors. Counselors who met criteria and chose to participate were consented into the study. The recruitment and training of counselors occurred in eight successive waves; the start of each wave was separated by 8 weeks. Counselors completed self-report instruments that collected demographic information (e.g., gender, ethnicity, educational level, age) and employment factors (e.g., number of years in the drug-counseling field, number of years at their current position, history of drug dependence, self-identified theoretical orientation) prior to the start of training. At each assessment time point (i.e., baseline, post-workshop, week 8, and week 20), counselors completed the HRQ (Miller et al., 1991). Counselor's audio recorded the assessment session with a different client each time, and no client was allowed to participate in more than one session. Counselors were not allowed to record sessions with clients on their caseload, so as to avoid possible coercion during client recruitment, and this ensured that all interviews were first time interviews with a new patient.

2.2.2. Workshop training

Counselors completed a 14-hour, 2-day MI training workshop led by a trainer who was a member of the Motivational Interviewing Network of Trainers (MINT) at either their treatment site or at the New York State Psychiatric Institute. The workshop focused on MI Spirit, specific MI skills (Open Questions, Affirmations, Reflections, and Summary Statements), and eliciting and recognizing change talk. Didactic instruction, experiential exercises, and role-playing were used to convey the concepts. During the week following the workshop training, counselors recorded a counseling session with a client in their clinic, and completed the HRQ. This served as the post-workshop assessment, which captured any changes in MI skill over the time of the workshop training.

2.2.3. Post-workshop supervision

Counselors in the Workshop group attended the initial 2-day workshop but did not receive post-workshop supervision.

Counselors who received post-workshop supervision (TCS or Tape) participated in five MI practice sessions with standard patients, portrayed by trained actors. The actor would travel to each counselor's place of employment, and simulate a clinical interaction based on a pre-determined role. The standardized patient cases are available from the first author (Smith et al., 2007) and were developed by the supervisors based upon conglomerations of actual cases. There are 7 different characters so that each role-play would be unique. Each character was presenting to the counselor as a new client with a substance use problem. The actors were working actors living in NYC and they varied in gender, race, and age. All actors were trained in all cases when they were first hired and then, every six weeks, all actors working on the project would come together for "refresher training". Actors had to be adept at improvisation while adhering to the specifics of the character. Audiotapes of the simulated clinical encounters with the standard patients (i.e., actors) were obtained for counselors in both the TCS and Tape conditions.

Supervisors listened to the audiotaped session, rated 20 min of the session using the MITI 2.0, rated the session for MITI Spirit and Empathy, and took notes on the counselor's MI style. All counselors received a graphical representation of their MI performance by email. The supervisor then called the counselor to review the graph, provide feedback, discuss the counselor's experience, and conduct role-plays to practice the use of MI skills. Each supervision session was conducted prior to the next practice interview.

Counselors in the TCS group received real time feedback via an earbud connected to the telephone line during the simulated clinical encounter (Smith et al., 2012). The TCS model was developed to provide immediate performance feedback that differentially reinforced MI-consistent and MI-inconsistent counseling styles, and allowed the supervisor to model MI Spirit and style. Counselors and supervisors discussed the session following its completion. In the tape group, supervisors rated a 20-min segment of the session, sent the counselor a graphical representation of MI performance, and reviewed the feedback during a phone call prior to the next TCS session.

2.3. Measures

2.3.1. Helpful Responses Questionnaire

The Helpful Responses Questionnaire (HRQ; Miller et al., 1991) is a 6-item, pen-and-paper measure of empathy—an essential MI skill—with "empathy" operationalized as the number and quality of reflections a counselor offers. This is an open-response questionnaire, which can be administered individually or in groups, can be given repeatedly to assess for change in ability, is not time consuming, and is easy to score.

The six items are statements by theoretical clients on various topics, such as excessive alcohol use or trouble with friends. Respondents are instructed "In each case write the *next* thing that you would say if you wanted to be helpful" (Miller et al., 1991; p. 444). Administration of the HRQ takes 15–20 min; scoring of the HRQ takes 10–15 min. Each

item is scored based on the quality of reflective listening statements contained in the counselor's written response, and the presence or absence of communication "roadblocks", i.e., responses that actively interrupt the flow of communication (Gordon, 1970). There are 12 roadblocks listed by Thomas Gordon, some examples are: 1. Ordering, directing, or commanding: this is when a direction is given with the force of some authority behind it; such as, "Don't say that." or "You've got to face up to reality". 2. Giving advice, making suggestions, providing solutions: this is when the individual draws on his/her own store of knowledge and experience to recommend a course of action; such as "Why don't you…" or "Have you tried…". 3. Interpreting or analyzing: seeking out the hidden meaning is very common and tempting to counselors; such as "You don't really mean that." or "Do you know what your real problem is?" Not all of the roadblocks are inconsistent with MI, but they do interrupt the flow of communication and force clients to address the counselor's thoughts rather than their own.

Each item on the HRQ is scored from 1 to 5, based on depth of reflection. A rating of 1 is assigned when the response contains no reflection, but does include at least one roadblock. A score of 2 is assigned when the response contains neither a reflection nor a roadblock, or when the response contains a reflection of any type and a roadblock. A response that contains a reflection that merely repeats the content already stated is scored 3. When the reflection adds inferred meaning that appears appropriate or plausible, a score of 4 is assigned. A score of 5 is given when the response includes either a reflection of feeling that fits the original statement, or an appropriate metaphor or simile. When multiple responses are made, the highest score is given—unless a roadblock is included, which would result in a score of 2 (Miller et al., 1991; p. 445). The score of each of the six items is totaled for the total HRQ score; scores can range from 6 to 30. The coefficient Cronbach's Alpha of this 6-item scale was 0.84 (Miller et al., 1991).

All participants completed the HRQ at each time point and participants were never told their score or what would make a "better" response.

Other research supports the use of the HRQ as a repeated measure for training. Hartzler, 2015, used a revised HRQ, the HRQ-CM, with clinicians in an opiate treatment program as they were trained in contingency management. In a limited sub-sample of clinicians, test-retest results showed the summary score was nominally reduced at re-test. This supports the notion that simply taking the test again will not lead to an increased score. At the conclusion of training, the HRQ-CM did detect significant training effects ($p < 0.001$). This further supports the notion that the HRQ is sensitive to training effects and not re-test effects.

A research assistant (RA) trained by the first author scored each HRQ. The lead author double coded 10% of the HRQs. The HRQs identified for double coding were randomly selected by a computer model selecting for an equal number of assessments at each time point. Double coding was completed independently and coders were blind to assessment point and training condition. A research staff member not affiliated with this study held the master list linking identifiers with patient identity and session number. ICC analyses comparing Coder 2 to Coder 1 were 0.79. This is "excellent" according to Cicchetti's (1994) guidelines for determining the utility of clinical instruments (<0.40 *poor*; 0.40 to 0.59 *fair*; 0.60 to 0.74 *good*; 0.75 to 1.00 *excellent*).

2.3.2. Motivational Interviewing Treatment Integrity (MITI 2.0; Moyers et al., 2005)

The MITI 2.0 was used to quantify a counselor's MI skillfulness during interactions with a client at each of the four assessment points. The MITI 2.0 is a coding system developed to rate the quality and quantity of counselor MI counseling behaviors during a randomly selected 20-minute audiotaped segment of a session. Counselors are rated along two global dimensions (Spirit and Empathy), and on the frequencies of seven counseling behaviors. The global score of Spirit is used to capture the degree to which the counselor supports the client's autonomy, evokes the client's reasons for change, and fosters a collaborative counseling environment. The global score of Empathy is used to capture the degree to which the counselor communicates understanding or makes an effort to understand the client's perspective. Global ratings are made along a continuum ranging from 1 (*low*) to 7 (*high*). The MITI 2.0 behavioral counts are based on specific types of counselor language that are important in MI. They are Open Questions, Closed Questions, Simple Reflections, Complex Reflections, Giving Information, MI Adherent (MIA), and MI Non-Adherent (MINA) statements. This differs from more recent versions of the MITI (e.g., MITI 3.0–4.2) in that "Spirit" is no longer calculated as one global score.

For the present investigation, the two MITI 2.0 global scores of Spirit and Empathy, three MITI 2.0 summary scores, and the MINA behavior count were used in the analyses. The three MITI 2.0 summary scores included were Percent Complex Reflections (the number of Complex Reflections made, divided by the total number of reflections made, multiplied by 100), Reflection to Question Ratio (the total number of reflections made, divided by the total number of Open and Closed Questions asked), and Percent MIA (the total number of MIA statements divided by MIA plus MINA statements, multiplied by 100). In addition, the number of MINA statements made by the counselor was counted—MINA statements are those that are not consistent with an MI style (e.g., labeling, arguing, giving advice without permission). Other MITI 2.0 scores were not included in the present analyses, as they assess skills outside the scope of the HRQ.

ICCs for the MITI 2.0 global scores and behavior counts were classified according to Cicchetti's (1994) guidelines for determining the utility of clinical instruments (<0.40 *poor*; 0.40 to 0.59 *fair*; 0.60 to 0.74 *good*; 0.75 to 1.00 *excellent*). ICCs (for a fixed set of raters; Shrout & Fleiss, 1979) for the MITI 2.0 behavioral indices ranged from fair to excellent: empathy 0.50; spirit 0.60; Simple Reflections 0.40; Complex Reflections 0.64; MI Non-Adherent (MINA) 0.72; MI Adherent (MIA) 0.47; open questions 0.74; and Reflection to Question Ratio 0.88. Percent agreement for global scores that were calculated were: Spirit: identical ratings from two blinded coders = 37%, ratings within one point from two blinded coders = 76%; Empathy: identical ratings from two blinded coders = 34%, ratings within one point from two blinded coders = 71%.

In general, MI training aims to guide the development of a counseling repertoire that is characterized by stronger MI Spirit (a collaborative alliance), a low frequency of MINA statements, a greater use of Complex Reflections (relative to Simple Reflections), and the use of more reflections and fewer questions (a greater Reflection to Question Ratio). Trained supervisors rated the audiotaped assessment sessions for MI proficiency using the MITI 2.0, and no supervisor rated a counselor that had been his or her supervisee during the study. All raters were blind to the supervision group assignment and timing of the session (i.e., baseline, post-workshop, week 8, or week 20).

3. Data analysis

Mean HRQ scores were calculated at four time-points and paired *t*-tests were used to test for mean changes over time. To test the first hypothesis, i.e., that a positive association would be demonstrated between the HRQ and MITI 2.0 scores at each assessment point, total HRQ score was correlated with six MITI 2.0 scores (i.e., Empathy, Spirit, Percent Complex Reflections, Reflection to Question Ratio, Percent MIA statements, and number of MINA statements) at each assessment time point (i.e., baseline, post-workshop, week 8, and week 20). Pearson correlations were used to assess the relation between HRQ score and the following MITI 2.0 indices: Spirit and Empathy, Percent MI Adherent Statements and Percent Complex Reflections. Spearman correlations were used to assess the relation between HRQ score and number of MI Non-Adherent statements and Reflection to Question Ratio, due to non-normally distributed data for MI Non-Adherent statements and Reflection to Question Ratio. Correlations were calculated at each of the four assessment points. A conservative approach was taken to protect

against Type I errors by employing a Bonferroni correction. Hypotheses were tested at a significance level of $p < 0.008$.

To evaluate the second hypothesis, i.e., that counselors receiving supervision would score higher on the HRQ than those who do not get supervision, mixed effects linear models were applied using a generalized estimating equation (GEE) analytical framework. The GEE analytical framework does not require complete outcome data across time points, so that cases missing data at one but not both outcome points could be included in the analyses under the assumption of missing at random. The linear model included contrast terms that tested the effect of supervisory condition (TCS versus Workshop; Tape versus Workshop; TCS versus Tape), time (week 8 versus week 20), and treatment clinic (outpatient drug-free versus residential; residential versus methadone maintenance; outpatient drug-free versus methadone maintenance), baseline HRQ score, and interaction between time and supervision condition. The generalized linear model (GLM) methodology was chosen to handle within-subject correlations (Diggle, Heggarty, Liang, & Zeger, 2002). PROC GENMOD in SAS 9.4 was used to conduct the analysis.

PROC LOGISTIC in SAS was used to calculate a ROC curve to analyze the utility of the HRQ in predicting MI proficiency. HRQ scores at post-workshop were tested for their ability to predict MI proficiency on the MITI 2.0. MI proficiency was defined as Reflection to Question Ratio ≥ 1.

4. Results

4.1. Participants

Of the 100 counselor-participants enrolled in the parent study, 97 completed the MI workshop and were randomized to one of the three supervisory conditions. Of the 97 counselors randomized, 32 were assigned to TCS, 32 were assigned to Tape, and 33 were assigned to Workshop (Smith et al., 2012). Demographics of participants can be found in Table 1, and are reported in the main outcomes article (Smith et al., 2012). Counselors represented a variety of educational levels, work experience, and therapeutic orientations.

4.2. Helpful Responses Questionnaire

Table 2 shows the means and standard deviations of all of the HRQ and MITI 2.0 scores. Scores on the HRQ can range from 6 to 30; higher scores are better, though the scale does not have designated cutoffs. Baseline HRQ scores ranged from 6 to 24 (mean = 10.5, SD = 3.6) and increased at post-workshop (mean = 17.6, range 6–24, SD = 4.6) ($t = 11.59$, $p < 0.01$). At week 8, HRQ scores ranged from 7 to 25 (mean = 18.5, SD = 4.2) and were not significantly different from post-workshop ($t = 1.32$, $p = 0.19$). At final follow-up (i.e., week 20), HRQ scores ranged from 6 to 24 (mean = 17.8, SD = 4.6) and were not significantly different from post-workshop ($t = 0.23$, $p = 0.82$).

4.2.1. Motivational Interviewing Treatment Integrity 2.0

Six MITI 2.0 scores were included in our analyses: Spirit, Empathy, Percent Complex Reflections, Question to Reflection Ratio, Percent MI Adherent Statements; and Number of MI Non-Adherent Statements. In general, MITI 2.0 scores improved from baseline to post- workshop, and again to week 8. Mean and standard deviations of MITI 2.0 scores by time point can be seen in Table 2. Full details on MITI 2.0 results can be found in Smith et al. (2012).

4.2.2. Hypothesis 1: HRQ and MITI 2.0 correlations

There were several significant correlations between HRQ total scores and MITI 2.0 subscales at each time-point after workshop, which are shown in Table 3. Though not present at baseline, the most consistent pattern was that HRQ score was significantly correlated with the MITI 2.0 Reflection to Question Ratio across post-workshop ($r = 0.33$; $p = 0.002$), week 8 ($r = 0.34$, $p = 0.002$) and week 20 ($r = 0.38$, $p < 0.001$). HRQ scores at baseline were lower and had less variability than at other time points thus, perhaps, limiting detection of potential correlations with the MITI 2.0. At week 8, HRQ score was significantly correlated with Empathy ($r = 0.30$, $p = 0.007$). At week 20, HRQ score was significantly correlated with Spirit and Empathy ($r = 0.32$, $p = 0.003$ for both). HRQ score did not significantly correlate with any other MITI 2.0 index at any time-point. Most of the correlations were in the small ($r = 0.10$) to medium ($r = 0.30$) range, as described by Cohen (1992).

Table 1
Demographic characteristics of the counselors (N = 97) randomly assigned to one of three conditions for training in MI: Teleconferencing Supervision (TCS), Tape-based supervision (Tape), and Workshop only (Workshop).

Variables	TCS ($n = 32$)	Tape ($n = 32$)	Workshop ($n = 33$)
Gender			
Male	41% ($n = 13$)	34% ($n = 11$)	30% ($n = 10$)
Female	59% ($n = 19$)	66% ($n = 21$)	70% ($n = 23$)
Age	45.9 (9.5)	42.9 (11.7)	42.9 (12.4)
Race			
African American	28% ($n = 9$)	44% ($n = 14$)	48% ($n = 16$)
Caucasian	34% ($n = 11$)	25% ($n = 8$)	27% ($n = 9$)
Hispanic	38% ($n = 12$)	19% ($n = 6$)	22% ($n = 7$)
Other	0% ($n = 0$)	12% ($n = 4$)	3% ($n = 1$)
Education			
High school	12% ($n = 4$)	9% ($n = 3$)	24% ($n = 8$)
College level	44% ($n = 14$)	41% ($n = 13$)	43% ($n = 14$)
Graduate degree	38% ($n = 12$)	47% ($n = 15$)	33% ($n = 11$)
Treatment clinic			
Drug-free outpatient	41% ($n = 13$)	31% ($n = 10$)	30% ($n = 10$)
Methadone maintenance	47% ($n = 15$)	50% ($n = 16$)	58% ($n = 19$)
Inpatient residential	12% ($n = 4$)	19% ($n = 6$)	12% ($n = 4$)
Years in the field	8.4 (6.3)	8.5 (8.1)	7.8 (6.3)
Years in current position	5.5 (5.3)	4.3 (4.6)	5.7 (5.5)
In recovery for a substance use problem	34% ($n = 11$)	41% ($n = 13$)	27% ($n = 9$)
Therapeutic orientation[a]			
AA/NA	39% ($n = 13$)	34% ($n = 11$)	25% ($n = 8$)
CBT	72% ($n = 24$)	81% ($n = 26$)	84% ($n = 27$)
MI	24% ($n = 8$)	22% ($n = 7$)	38% ($n = 12$)
Harm reduction	51% ($n = 17$)	38% ($n = 12$)	34% ($n = 11$)

[a] Categories are not mutually exclusive.

Table 2
HRQ and MITI 2.0 means and standard deviations (N = 97[a]).

Variable	Baseline Mean (SD)	Post-workshop Mean (SD)	Week 8 Mean (SD)	Week 20 Mean (SD)
HRQ	10.54 (3.59)	17.64 (4.58)	18.54 (4.18)	17.80 (4.64)
Spirit	4.11 (1.04)	4.80 (1.21)	4.78 (1.09)	4.89 (1.12)
Empathy	4.30 (1.12)	4.80 (1.15)	4.81 (1.03)	4.85 (1.05)
MI Non-Adherent	2.54 (3.72)	1.07 (2.53)	1.06 (2.16)	1.06 (2.33)
Percent MI Adherent	48.42 (36.47)	70.76 (39.17)	71.40 (39.08)	77.13 (33.42)
Percent Complex Reflection	36.99 (26.01)	45.71 (22.73)	44.64 (22.68)	44.81 (25.64)
Reflection to Question Ratio	0.47 (0.36)	0.78 (0.75)	0.82 (0.70)	0.74 (0.77)

Notes: HRQ = Helpful Responses Questionnaire; MITI = Motivational Interviewing Treatment Integrity.
[a] Missing 7–17 cases for HRQ, and 5–10 cases for MITI during these four time points.

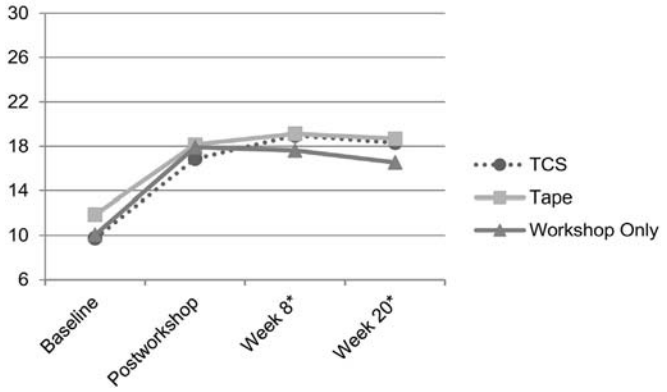

Fig. 1. HRQ scores across time by Training Group. HRQ scores of counselors randomized to supervision conditions: Teleconferencing Supervision (TCS), Taped-based supervision (Tape), or Workshop only. Time is at baseline (N = 90); post-workshop (N = 90); Week 8 (after supervision; N = 80), and Week 20 (N = 89). *Counselors in TCS scored significantly higher than those in Workshop only at Week 8 and Week 20. There was no significant difference between counselors in TCS and Tape.

4.2.3. Hypothesis 2: treatment effects on HRQ scores

Inspection of Fig. 1 shows small differences in HRQ scores between the three training groups over time (week 8 vs. week 20). The HRQ scores did differ significantly by supervisory condition ($X^2(2) = 7.88$, $p = 0.02$), and this effect did not vary by time-points ($X^2(1) = 0.21$, $p = 0.90$). Counselor-participants in the Workshop group (i.e., no supervision) scored lower than the TCS group at week 8 and week 20 ($z = 3.10$, $p = 0.002$). HRQ scores were not significantly different between the TCS and Tape supervision groups at any time point. All analyses controlled for baseline scores.

4.2.4. Additional analysis: HRQ scores predict MI proficiency

By using the threshold for proficiency on MITI 2.0 Reflection to Question Ratio (R:Q) as the gold standard, a ROC curve was estimated to determine the optimum cutoff level of HRQ score at post-workshop. R:Q is considered an advanced MI skill and it was the MITI 2.0 score most often significantly correlated with the HRQ. The sensitivity (i.e., true positive rate) and specificity (i.e., true negative rate) for HRQ score 21 were 44% and 78% respectively (Kappa 0.25), see Table 4. Other HRQ cut-off scores did not yield better results and are also presented in Table 4.

5. Discussion

This secondary analysis of data evaluating relationships between the written analog HRQ, a pen-and-paper test of empathy, and the independent coding of audiotaped sessions system, MITI 2.0, found that the HRQ correlated with MITI 2.0 Reflection to Question Ratio, Spirit, and Empathy. Expected correlations between HRQ and other MITI 2.0 subscales were not found. The first hypothesis, i.e., that a positive association would be demonstrated between HRQ and MITI 2.0 scores at each assessment point, was partially supported. HRQ correlated with some but not all MITI 2.0 measures: the Reflection to Question Ratio (R/Q) scores, and the global scores of Spirit and Empathy after supervision. The size of the significant correlations was in the medium range (Cohen, 1992). As we proposed in our second hypothesis, the HRQ was sensitive to training effects, with counselors scoring higher after receiving supervision, compared to those who did not receive supervision. Taken together, these data suggest that the HRQ has promise as a quick assessment of progress in MI training, though with limitations and room for improvement.

A ROC curve analysis was completed to assess the utility of the HRQ in predicting MI skill. For this analysis, MITI 2.0 Reflection to Question Ratio < 1 = not-proficient; ≥1 = proficient. The HRQ score was shown to have limited sensitivity and specificity. The "sensitivity" is the fraction of people with the criteria that the test correctly identifies as positive. The "specificity" is the fraction of people without the criteria that the test correctly identifies as negative. At the HRQ cutoff score of 21, the HRQ would correctly identify 78% counselors as "MI not-proficient" (i.e. specificity) but miss-identify 22% of counselors who are proficient.

HRQ was associated with MITI 2.0 indices on reflections and empathy, but not other scores. This result would support the intended purpose of the HRQ, i.e., "a reliable and valid pen-and-paper measure of accurate empathy or active listening" (Miller et al., 1991). It is possible that this pattern of correlations reflects an evolution in learning MI, which is likely to begin with the skill of generating more reflections than questions. This finding is consistent with Miller and Moyers' (2006) proposed stages in learning MI: according to their hypothesized sequence, generating accurate reflections and showing MI spirit are two of the earliest skills counselors acquire. HRQ-like items could be developed to tap more advanced MI microskills, which are acquired as MI skill increases with more training and experience (e.g., eliciting and differentially reinforcing change talk). Related to the issue of the stages in learning MI, the need to tailor training to counselors' current stage of learning MI has been emphasized in the literature (Miller & Moyers, 2006; Barwick et al., 2012; Forsberg et al., 2011). Another interesting result in this study was that significant correlations were found after supervision and not at baseline, which does not support the idea of

Table 3
Correlations between HRQ and MITI 2.0.

Variable	Baseline correlation with HRQ ($n = 80$)	Post-workshop correlation with HRQ ($n = 85$)	Week 8 correlation with HRQ ($n = 78$)	Week 20 correlation with HRQ ($n = 84$)
Spirit	$r = 0.18$, $p = 0.11$	$r = 0.06$, $p = 0.56$	$r = 0.25$, $p = 0.03$	**$r = 0.32$, $p < 0.01$**
Empathy	$r = 0.14$, $p = 0.22$	$r = 0.18$, $p = 0.11$	**$r = 0.30$, $p < 0.01$**	**$r = 0.32$, $p < 0.01$**
MI Non-Adherent	$r = -0.13$, $p = 0.24$	$r = -0.18$, $p = 0.10$	$r = 0.03$, $p = 0.81$	$r = 0.03$, $p = 0.80$
Percent MI Adherent	$r = 0.16$, $p = 0.18$	$r = 0.12$, $p = 0.33$	$r = -0.05$, $p = 0.70$	$r = -0.01$, $p = 0.95$
Percent Complex Reflection	$r = 0.20$, $p = 0.08$	$r = -0.00$, $p = 0.99$	$r = -0.07$, $p = 0.56$	$r = 0.17$, $p = 0.12$
Reflection to Question Ratio	$r = 0.17$, $p = 0.13$	**$r = 0.33$, $p < 0.01$**	**$r = 0.34$, $p < 0.01$**	**$r = 0.38$, $p < 0.01$**

Notes: HRQ = Helpful Responses Questionnaire; MITI = Motivational Interviewing Treatment Integrity 2.0.

Table 4
Receiver Operating Characteristic (ROC) curve analysis for Helpful Responses Questionnaire (HRQ) for cut scores 17–23.

	Sensitivity (true positives) proficient on R:Q (>1)	False negatives	Specificity (true negatives) not proficient on R:Q (<1)	False positives
HRQ				
17	1.00 (=18/18)	0.00 (=0/18)	0.40 (=27/67)	0.60 (=40/67)
18	0.94 (=17/18)	0.06 (=1/18)	0.43 (=29/67)	0.57 (=38/67)
19	0.67 (=12/18)	0.33 (=6/18)	0.55 (=37/67)	0.45 (=30/67)
20	0.56 (=10/18)	0.44 (=8/18)	0.66 (=44/67)	0.34 (=23/67)
21[a]	0.44 (=8/18)	0.56 (=10/18)	0.78 (=52/67)	0.22 (=15/67)
22	0.39 (=7/18)	0.61 (=11/18)	0.85 (=57/67)	0.15 (=10/67)
23	0.17 (=3/18)	0.83 (=15/18)	0.90 (=60/67)	0.10 (=7/67)

[a] Using the cut score of 21, the HRQ was able to correctly identify 78% of participants who were not proficient on the MITI 2.0 Reflection to Question Ratio.

using the HRQ to differentiate between practitioners who come into training with a more MI-consistent counseling approach from those who lack an MI-consistent style and are thus in need of more training. Nonetheless, some authors have underscored the need to screen counselors for baseline fundamental MI skills, like empathy, so this appears as a topic deserving of further investigation (Gaume, Gmel, Faouzi, & Daeppen, 2009; Moyers & Miller, 2013).

Prior studies examined the HRQ's ability to measure counselor empathy. Baer et al. (2012) found significant correlations between the HRQ and the CASPI, while Decker et al. (2014) found no significant correlations with the TES at any assessment point and only a small positive correlation when combining data across time points. A possible explanation is that HRQ measures only one aspect of empathy – the cognitive - while the TES measures all the aspects the authors of the scale hypothesize as the different facets of empathy (i.e., affective, attitudinal, setting aside one's view to enter client's world, and attunement to momentary changes in the client's presentation - in addition to the cognitive aspect). Barwick et al. (2012) noted that diverse measures of behavior change are not necessarily equivalent and highlighted the need to consider such differences when comparing measures of MI competence. Our results lining up more closely with Baer et al. (2012) regarding the CASPI, could stem from the fact that the CASPI uses audio vignettes very similar to the written ones in the HRQ, to which counselors are asked to respond. The HRQ is narrower in scope as a measure of MI skill compared to an interview-based instrument like the MITI; the HRQ is a written measure while the MITI rates verbal behavior during sessions. MITI scores represent interactions with actual clients while HRQ scores represent reactions to a fictional client.

Also, a possible critical difference between the HRQ and our use of the MITI 2.0 in this study is that the HRQ posits six scenarios in response to which the counselor could provide a helpful statement, but only two of the six circumstances presented on the HRQ involve any alcohol use, and none involve any drug use (e.g., HRQ vignette 2 presents the case of a 36 year old client complaining about a nosy neighbor who is always borrowing things that he never returns, and phones late at night). In contrast, when recording the MI sessions, counselors were instructed to speak with the patient about their substance use behavior, so while both the HRQ and MITI 2.0 may be coding empathy, the two measures are eliciting empathic statements for different types of problems, under different circumstances.

An important drawback of measures that involve audiotaping of sessions, widely reported in the literature, has been counselors' variable adherence with audiotape submission requirements (Barwick et al., 2012; Bennett, Moore, et al., 2007; Bennett, Roberts, et al., 2007; Madson & Campbell, 2006; Vachon et al., 1995). When a counselor is pressed for time and leery of training, the HRQ may be more likely to be completed and submitted for review. The HRQ is substantially easier, faster, less vulnerable to technical glitches than recording equipment, less intrusive, and may be experienced as less threatening to counselors than other measures of MI skill.

The strengths of this study included a notable sample size, which afforded adequate power to detect a meaningful effect, a longitudinal training design, and relatively little missing data for a study of this type. The HRQ coders were blind to participant identity, randomization, or session number. All MITI 2.0 raters and the trained supervisors were also blinded to participant identity, randomization, and session number. Limitations of the study include the fact that all participants completed the HRQ four times thus raising concerns of test-retest effects. Thus, we could not isolate the effects of practice. However, other research supports the use of the HRQ as a repeated measure for training. Hartzler, 2015, used a revised HRQ, the HRQ-CM, with clinicians in an opiate treatment program as they were trained in contingency management. In a limited sub-sample of clinicians, test-retest results showed the summary score was nominally reduced at re-test. This supports the notion that simply taking the test again will not lead to an increased score. At the conclusion of training, the HRQ-CM did detect significant training effects ($p < 0.001$). This further supports the notion that the HRQ is sensitive to training effects and not re-test effects. Thus any changes seen in HRQ scores over time, could be representative of a change in counseling style. Future work could rectify these limitations by varying groups and re-coding.

In summary, these findings suggest that the HRQ may not be useful as a screening tool to identify counselors with higher vs. lower baseline proficiency in MI, but may provide a shorthand measure of reflective listening skill during training. Thus, the HRQ can be used as a quick assessment of progress in MI training. More work is needed to both develop vignettes relevant to substance use for the HRQ, and to create a more comprehensive pen-and-paper measure of MI skill that will be easy to collect from trainees and less time intensive than coding MITI assessments.

Research support

NIDA Grant DA16950; K23 DA 021850 Carpenter; K24 DA 022412 Nunes.

References

Baer, J. S., Carpenter, K. M., Beadnell, B., Stoner, S. A., Ingalsbe, M. H., Hartzler, B., et al. (2012). Computer assessment of simulated patient interviews (CASPI): Psychometric properties of a web-based system for the assessment of motivational interviewing skills. *Journal of Studies on Alcohol and Drugs*, *73*(1), 154–164.

Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*, *73*(1), 99–106.

Barwick, M. A., Bennett, L. M., Johnson, S. N., McGowan, J., & Moore, J. E. (2012). Training health and mental health professionals in motivational interviewing: A systematic review. *Children and Youth Services Review*, *34*, 1785–1795.

Bennett, G. A., Moore, J., Vaughan, T., Rouse, L., Gibbins, J. A., Thomas, P., et al. (2007). Strengthening motivational interviewing skills following initial training: A randomised trial of workplace-based reflective practice. *Addictive Behaviors*, *32*(12), 2963–2975.

Bennett, G. A., Roberts, H. A., Vaughan, T. E., Gibbins, J. A., & Rouse, L. (2007). Evaluating a method of assessing competence in motivational interviewing: A study using simulated patients in the United Kingdom. *Addictive Behaviors*, *32*(1), 69–79.

Carroll, K. M., Nich, C., & Rounsaville, B. J. (1997). Contribution of the therapeutic alliance to outcome in active versus control psychotherapies. *Journal of Consulting and Clinical Psychology*, *65*(3), 510–514.

Childers, J. W., Bost, J. E., Kraemer, K. L., Cluss, P. A., Spagnoletti, C. L., Gonzaga, A. M., & Arnold, R. M. (2012, Nov). Giving residents tools to talk about behavior change: A motivational interviewing curriculum description and evaluation. *Patient Education and Counseling*, *89*(2), 281–287.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cohen, C. (1992). A power primer. *Psychological Bulletin*, *12*(1), 155–159.

Cucciare, M. A., Ketroser, N., Wilbourne, P., Midboe, A. M., Cronkite, R., Berg-Smith, S. M., & Chardos, J. (2012). Teaching motivational interviewing to primary care staff in the Veterans Health Administration. *Journal of General Internal Medicine*, *27*(8), 953–961.

Daeppen, J. B., Fortini, C., Bertholet, N., Bonvin, R., Berney, A., Michaud, P. A., ... Gaume, J. (2012). Training medical students to conduct motivational interviewing: A randomized controlled trial. *Patient Education and Counseling*, *87*(3), 313–318.

Decker, S. E., & Martino, S. (2013). Unintended effects on clinician's interest, confidence, and commitment in using motivational interviewing. *Drug and Alcohol Dependence*, *132*(3), 681–687.

Decker, S. E., Carroll, K. M., Nich, Canning-Ball, M., & Martino, S. (2013). Correspondence of motivational interviewing adherence and competence ratings in real and role-played client sessions. *Psychological Assessment*, *25*, 306–312.

Decker, S. E., Nich, C., Carroll, K. M., & Martino, S. (2014). Development of the therapist empathy scale. *Behavioural and Cognitive Psychotherapy*, *42*, 339–354.

Diggle, P., Heggarty, P., Liang, K. -Y., & Zeger, S. (2002). Analysis of longitudinal data (second edition) ISBN 0198524846; 396 pages; Oxford University Press; http://www.oup.co.uk/isbn/0-19-852484-6. *Pharmaceutical Statistics*, *3*, 147–148 (Doi: 101002/pst).

Doran, N., Hohman, M., & Koutsenok, I. (2011). Linking basic and advanced motivational interviewing training outcomes for juvenile correctional staff in California. *Journal of Psychoactive Drugs*, *43*(Suppl. 1), 19–26.

Forsberg, L., Forsberg, L. G., Lindqvist, H., & Helgason, A. R. (2010). Clinician acquisition and retention of motivational interviewing skills: A two-and-a-half-year exploratory study. *Substance Abuse Treatment, Prevention, and Policy*, *5*, 8-597X-5-8.

Forsberg, L., Ernst, D., & Farbring, C. A. (2011). Learning motivational interviewing in a real-life setting: A randomized controllled trial in the Swedish prison service. *Criminal Behavior and Mental Health*, *21*, 177–188.

Gaume, J., Gmel, G., Faouzi, M., & Daeppen, J. B. (2009). Counselor skill influences outcomes of brief motivational interventions. *Journal of Substance Abuse Treatment*, *37*(2), 151–159.

Gordon, T. (1970). A theory of healthy relationships and a program of parent effectiveness training. *New directions in client-centered therapy* (pp. 407–425).

Hartzler, B. (2015). Adapting the Helpful Responses Questionnaire to assess communication skills involved in delivering contingency management: Preliminary psychometrics. *Journal of Substance Abuse Treatment*, *55*, 52–57.

Klonek, F. E., & Kauffeld, S. (2015). Providing engineers with OARS and EARS: Effects of a skills-based vocational training for engineers in higher education. *Higher Education, Skills and Work-Based Learning*, *5*(2), 117–134.

Madson, M. B., & Campbell, T. C. (2006). Measures of fidelity in motivational enhancement: A systematic review. *Journal of Substance Abuse Treatment*, *31*(1), 67–73.

Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carroll, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *19*(2), 181–193.

Martino, S., Gallon, S., Ball, S. A., & Carroll, K. M. (2008). A step forward in teaching addiction counselors how to supervise Motivational Interviewing using a clinical trials approach. *Journal of Teaching in the Addictions*, *6*(2), 39–67. https://doi.org/10.1080/15332700802127946.

Miller, W. R., & Rollnick, S. (2013). *Motivational Interviewing* (3rd Edition). New York: Guilford.

Miller, W. R., Hedrick, K. E., & Orlofsky, D. R. (1991). The Helpful Responses Questionnaire: A procedure for measuring therapeutic empathy. *Journal of Clinical Psychology*, *47*(3), 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*(04), 457–471.

Miller, W. R., & Moyers, T. B. (2006). Eight stages in learning motivational interviewing. *Journal of Teaching in the Addictions*, *5*(1), 3–17.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist*, *64*(6), 527.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, *72*(6), 1050–1062.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, *28*(1), 19–26.

Moyers, T. B., & Miller, W. B. (2013). Is low therapist empathy toxic? *Psychology of Addictive Behaviors*, *27*(3), 878–884.

Olmstead, T., Carroll, K. M., Canning-Ball, M., & Martino, S. (2011). Cost and cost-effectiveness of three strategies for training clinicians in motivational interviewing. *Drug and Alcohol Dependence*, *116*(1–3), 195–202.

Rosengren, D. B., Hartzler, B., Baer, J. S., Wells, E. A., & Dunn, C. W. (2008). The video assessment of simulated encounters-revised (VASE-R): Reliability and validity of a revised measure of motivational interviewing skills. *Drug and Alcohol Dependence*, *97*(1–2), 130–138.

deRoten, Y., Zimmermann, G., Ortega, D., & Despland, J. N. (2013). Meta-analysis of the effects of MI training on clinicians' behavior. *Journal of Substance Abuse Treatment*, *45*(2), 155–162.

Shafer, M. S., Rode, R., & Chong, J. (2004). Using distance education to promote the transfer of motivational interviewing skills among behavioral health professionals. *Journal of Substance Abuse Treatment*, *26*(2), 141–148.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Smith, J. L., Brooks, A. C., Carpenter, K. M., Levin, D., Schreiber, E. A., Travaglini, L. A., ... Nunes, E. V. (2007). *Standard role plays for use in clinician training.* Available from the first author at smitjen@nyspi.columbia.edu.

Smith, J. L., Carpenter, K. M., Amrhein, P. C., Brooks, A. C., Levin, D., Schreiber, E. A., et al. (2012). Training substance abuse clinicians in motivational interviewing using live supervision via teleconferencing. *Journal of Consulting and Clinical Psychology*, *80*(3), 450–464.

Stein, L. A. R., Clair, M., Soenksen-Bassett, S., Martin, R. A., & Clarke, J. G. (2015). Studying process and proximal outcomes of supervision for motivational interviewing. *Training and Education in Professional Psychology*, *9*(2), 175.

Vachon, D. O., Susman, M., Wynne, M. E., Birringer, J., Olshefsky, L., & Cox, K. (1995). Reasons therapists give for refusing to participate in psychotherapy process research. *Journal of Counseling Psychology*, *42*(3), 380.

Wain, R. M., Kutner, B. A., Smith, J. L., Carpenter, K. M., Hu, M. C., Amrhein, P. C., & Nunes, E. V. (2015). Self-report after randomly assigned supervision does not predict ability to practice Motivational Interviewing. *Journal of Substance Abuse Treatment*, *57*, 96–101.

Walters, S. T., Matson, S. A., Baer, J. S., & Ziedonis, D. M. (2005). Effectiveness of workshop training for psychosocial addiction treatments: A systematic review. *Journal of Substance Abuse Treatment*, *29*(4), 283–293.