9-1-2005

Development of the Motivational Interviewing Supervision and Training Scale

Michael B. Madson
Marquette University

Todd Campbell
Marquette University, todd.campbell@marquette.edu

David E. Barrett
Center for Addiction and Behavioral Health Research

Michael J. Brondino
Center for Addiction and Behavioral Health Research

Timothy P. Melchert
Marquette University, timothy.melchert@marquette.edu

Accepted version. *Psychology of Addictive Behaviors*, Vol. 19, No. 3 (September 2005): 303-310. DOI.

Development of the Motivational Interviewing Supervision and Training Scale

Michael B. Madson
Department of Counseling and Educational Psychology
Marquette University
Center for Addiction and Behavioral Health Research
Milwaukee, WI
Todd C. Campbell
Department of Counseling and Educational Psychology
Marquette University
Center for Addiction and Behavioral Health Research
Milwaukee, WI
David E. Barrett
Center for Addiction and Behavioral Health Research
Milwaukee, WI
Michael J. Brondino
Center for Addiction and Behavioral Health Research
Milwaukee, WI
Timothy P. Melchert
Department of Counseling and Educational Psychology
Marquette University
Center for Addiction and Behavioral Health Research
Milwaukee, WI

[Citation: Journal/Monograph Title, Vol. XX, No. X (yyyy): pg. XX-XX. DOI.

The movement to use empirically supported treatments has increased the need for researchers and supervisors to evaluate therapists' adherence to and the quality with which they implement those interventions. Few empirically supported approaches exist for providing these types of evaluations. This is also true for motivational interviewing, an empirically supported intervention important in the addictions field. This study describes the development and psychometric evaluation of the Motivational Interviewing Supervision and Training Scale (MISTS), a measure intended for use in training and supervising therapists implementing motivational interviewing. Satisfactory interrater reliability was found (generalizability coefficient $p^2 = .79$), and evidence was found supporting the convergent and discriminant validity of the MISTS. Recommendations for refinement of the measure and future research are discussed.

The movement to use empirically supported treatments (ESTs) in both substance abuse and mental health settings continues to gain momentum. ESTs are often considered preferred treatments for a variety of psychological disorders because evidence of their efficacy has been demonstrated through randomized clinical trials (Waehler, Kalodner, Wampold, & Lichtenberg, 2000). This movement is not without controversy, however, as both researchers and clinicians question the methods with which ESTs are identified as well as how well ESTs generalize to practice settings (Davison, 1998; Waehler et al., 2000). Nonetheless, encouragement to use ESTs in psychological practice continues to grow.

This movement to increase the use of ESTs also raises questions regarding the strength of the empirical evidence regarding how well practitioners and researchers are implementing particular ESTs in practice as well as research settings. Carroll and colleagues (2002) suggested that a major challenge faced in technology transfer involves developing protocols that address both internal and external validity and ensuring that treatments are implemented as intended in terms of both adherence to the model as well as quality of the intervention. The present study describes the development of an instrument that can be used to address some of those questions within the context of motivational interviewing (MI), an increasingly popular form of therapy that is widely used in the addictions field.

Motivational interviewing is a directive, client-centered approach for eliciting behavior change by helping clients explore and resolve ambivalence (Miller & Rollnick, 2002). Motivational interviewing has

been studied fairly extensively and shows promise as an efficacious intervention in a variety of settings, including outpatient and residential treatment, medical settings, and employee assistance programs. Because of small sample sizes, convenience and homogeneous samples, and lack of randomization, however, the results of many of these studies need to be interpreted cautiously (Burke, Arkowitz, & Menchola, 2003; Dunn, Deroo, & Rivara, 2001).

Another major concern with several of the studies examining the efficacy of MI is the fidelity and quality with which the intervention was implemented. As seen in Table 1, there is variability in the degree to which studies of MI have described the training, supervision, and monitoring of therapists implementing MI. This is problematic given the concerns of some researchers that MI is sometimes implemented in a fashion that violates the spirit of the approach (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003; Rollnick & Miller, 1995). Miller (2001) also advised that new studies assess whether MI is being implemented faithfully and that this assessment must be accomplished through the direct monitoring of the intervention, as opposed to clinician self-report (see also Carroll et al., 2002).

The need to measure adherence to and quality of an implementation also applies in training and supervision contexts. When trainees and supervisees are learning to implement ESTs, supervisors need to evaluate whether they are adhering to the specific type of therapy that is being taught and the level of skill with which they are implementing the intervention, as providing evaluation and feedback in particular is viewed as a defining characteristic of supervision (Bernard & Goodyear, 1998). Despite the importance of evaluation in training and supervision, little research has examined the processes that are used to evaluate supervisees. Specifically, current methods of evaluation tend to be grounded in professional experience rather than empirically based evidence (Bernard & Goodyear, 1998). This practice continues despite the need for more rigorous evaluation procedures resulting from the increasing emphasis on accountability in health care fields and the continued development of therapeutic modalities that require highly developed skills. To address these needs, Bernard and Goodyear (1998) suggested that supervisors make clear distinctions between *formative* (i.e., process) and *summative* (i.e., outcome) evaluation, ensure that formative evaluations inform summative

evaluations, and monitor trainees' behavior to check the fidelity of the implementation of the intervention without merely monitoring adherence by ensuring that quality care is provided.

Currently, there are few measures that supervisors can use to provide objective feedback to supervisees implementing ESTs. With regard to MI specifically, we were able to locate no instrument that evaluated therapist adherence to this form of therapy, or the quality with which it is implemented, and that also had clinical utility as well as acceptable psychometric properties. A literature search found two measures that focus on these issues. Miller (2000) developed the Motivational Interviewing Skill Code (MISC) for rating therapist-client interactions in audio- or videotaped MI sessions to assess adherence to the MI approach. The reliability and validity of the data produced by the MISC vary considerably by item. For example, Tappin et al. (2000) found that intraclass correlation coefficients of interrater reliability for the global items on the MISC ranged from .39 to .53, and Moyers et al. (2003) found that these coefficients ranged from .25 to .86 for the MISC global items and from .00 to 1.00 for the behavioral counts. The construct validity of the MISC was examined by Miller and Mount (2001), who did find that therapy sessions by four MI experts were rated highly on the MISC. The MISC has reduced clinical utility, however, in that it is quite complex to learn and use and can take up to 4 hr to rate one therapy session (Tappin et al., 2000). Barsky and Coleman (2001) developed the Motivational Interviewing Process Code to assist training in MI by evaluating functional and dysfunctional MI skills, but evidence regarding its reliability and validity also is variable. Interrater agreement was found to be 51% and 75% for the instrument's two subscales. Also, it appears that although the measure was intended to evaluate skill acquisition, it assesses adherence more than quality of MI. In response to these concerns, and because of the need to consider clinical utility, we decided to develop an alternative measure that assesses information similar to that assessed by the MISC but that is targeted for use in either a clinical or research setting.

Method

Tapes and Therapists

A sample of 50 audiotaped sessions was randomly selected from 89 audiotaped therapy sessions that were submitted for fidelity monitoring and clinical supervision by four therapists participating in Project REFER (Referring Early for Early Recovery), a study conducted by the University of Connecticut through a grant from the Center for Substance Abuse Treatment that examined the effectiveness and cost-effectiveness of an adaptation of MI called Motivational Enhancement Therapy (Miller, Zweben, DiClemente, & Rychtarik, 1995) as a stand-alone intervention in outpatient addiction service settings. The four therapists (one man and three women) all had advanced professional degrees (one doctorate and three master's degrees) and an average of 16.5 years of clinical experience, although none had significant experience with MI prior to participating in this study.

All four therapists were trained by a team of research practitioners with considerable experience using MI. The training was based on the treatment manual used in the study (Barrett, Rugg, Zweben, Campbell, & Madson, 1998), and individual clinical supervision was provided on a regular basis. Because the three clinical sites in this study were in geographically separate locations, however, supervision and consultation were conducted by telephone.

Raters

The raters in this study were advanced doctoral students in counseling psychology. They included two women and one man, ranging in age from 25 to 30 years. All raters held master's degrees in a mental health related field and had from 2 to 5 years of postmaster's therapy experience. All three also had training and experience providing clinical supervision. The raters were relatively inexperienced in the use of MI; two had received training in MI, but only the third had used MI in therapy.

The raters were trained in the Motivational Interviewing Supervision and Training Scale (MISTS) using a modified version of the

procedure used by Carroll and colleagues (2000) in which raters attended one 2-hr and two 1.5-hr sessions that were facilitated by Michael B. Madson. The first session included a review of the purposes of the study, the key concepts of MI, the training manual and the rating scales, and the definitions and applications of key terms as defined in the manual. The second session reviewed and expanded on these objectives and included a practice session in which raters reviewed segments of audiotapes together (training tapes were not part of the 50 sample tapes), made ratings using the study measures and discussed how they made their rating decisions until 80% interrater agreement was achieved, discussed discrepancies and problems, and asked questions. The final session reviewed the above objectives and the specific plan for rating the tapes.

Measures

MISTS

The MISTS was designed to assist in the training and supervision of therapists implementing treatments using MI as a core element of the intervention. The principles and skills involved in this style of therapy are referred to as the "spirit of MI" and involve rolling with resistance, addressing ambivalence, and supporting client self-efficacy (Miller & Rollnick, 2002). The instrument is designed to provide a behavioral count of skills consistent with MI as well as assess the quality with which the intervention is delivered.

The MISTS includes two components: (a) behavioral count of the types of therapist responses uttered during sessions and (b) a 16-item global rating of the quality, MI fidelity, and effectiveness of therapist interventions. To complete the first component, a rater reviews a recorded therapy session and classifies each therapist utterance using broadly defined categories: open question, closed question, simple reflection, complex reflection, affirmation, summarization, interpretation, or providing information or advice. These categories are based on the therapist responses described by Miller and Rollnick (2002) as central to the appropriate implementation of MI (both common and specific components). This behavioral count section also provides the rater an opportunity to identify and classify a

missed opportunity if the therapist fails to elicit or reinforce client *change talk*, another primary goal of MI therapy (Miller & Rollnick, 2002). Obtaining a frequency count of the types of verbal behaviors a therapist uses in a session can be very helpful for training and supervision purposes and can be very helpful in assigning global ratings, the second component of the MISTS. Raters were able to complete the behavioral count in real time, and so the time to complete this component corresponded to the actual length of the recorded session.

The second component of the MISTS is completed after the behavioral counts of therapist responses are completed. This section involves making global ratings of aspects of MI therapy considered central to the approach and takes from 1 to 4 min to complete. An initial 27-item version of this component of the instrument was developed by a group consisting of three researchers and two practitioners, all with extensive experience in training and implementing MI. Initial items were generated on the basis of team discussions, review of the MISC, and literature on MI (Miller & Rollnick, 2002; Substance Abuse and Mental Health Services Administration, 1999) and core counseling skills (Hill & O'Brien, 1999; Ivey & Bradford Ivey, 2003). Discussions focused on item content and how these items fit with MI concepts. The initial version was tested in a pilot study conducted with the purpose of informing the research group about the process of using the measure and its utility, structure, and content. Two members from the research team reviewed 30 MI audiotapes from a separate study and provided feedback regarding the clarity of items and suggestions for improving the measure. As a result, behavioral anchors were developed for each item, and several narrowly focused items were combined into more general items (e.g., 5 items assessing types of therapist reflection responses were combined into 2 items focusing on simple reflections and complex reflections). Items were accepted when a group consensus was reached. In developing behavioral anchors, the development group followed the same procedure as for item generation, and a behavioral anchor was accepted when a group consensus was reached. After these changes were made, 16 items were retained for the final version of the MISTS, and a rating manual was developed. This final version of the MISTS was also reviewed for content and structure by five researchers independent of the developmental process.

The 16 items on the global rating section of the MISTS are organized into three categories: (a) specific active listening skills, (b) specific skills that demonstrate the spirit of MI, and (c) overall therapist ratings. Ratings are made on a 7-point Likert-type scale with behaviorally defined anchors at Points 1, 4, and 7 for each item, with low scores representing poor use of the skills being assessed. The total MISTS score is calculated by adding the score from each of the 16 individual items. For example, the behavioral anchors for the affirming item are as follows: 1 = *little or no attempt to identify client strengths or successes*; 4 = *maintains a nonjudgmental, accepting stance toward client goals and activities but little active affirming*; 7 = *regularly and systematically elicits and reinforces strengths, communicating a sense of optimism and hope*.

The five items in the *specific active listening skills* category include therapist use of questions (both open and closed), simple reflection (e.g., paraphrase, restatements), complex reflections (e.g., reflection with a twist, double-sided reflection), affirming, and summarizing. There are six items pertaining to the *spirit of MI* that reflect the underlying principles of the approach (Miller & Rollnick, 2002). These include engaging the client in the intervention process, eliciting and reinforcing client change talk, addressing client ambivalence, rolling with resistance, collaborating with the client, and supporting client self-efficacy. The five items in *overall therapist ratings* involve general evaluations of the use of active listening skills (i.e., questions, simple reflections, complex reflections, summary), appropriate sequencing of skills (i.e., open questions, affirmation, reflection, and summary), use of the spirit of MI (i.e., avoiding arguing, eliciting and reinforcing change talk), general response of the client (e.g., disengaged, argumentative), and the general effectiveness of the therapist in using MI.

Yale Adherence and Competence Scale (YACS; Corvino et al., 2000)

The YACS is a 50-item measure that evaluates general interventions common among most therapies as well as interventions associated with specific therapy modalities. The instrument includes six subscales. Three of the subscales (Assessment, General Support, and Goals for Treatment) assess the general interventions common

across therapies, and the other three subscales (Clinical Management, Twelve-Step Facilitation, and Cognitive-Behavioral Management) assess interventions specific to different psychotherapy modalities. For each item, raters judge both adherence to and quality of implementation. Frequency ratings range from 1 = *not at all* to 7 = *extensive*, and ratings of quality range from 1 = *very poor (therapist handled this in an unacceptable even toxic manner*) to 7 = *excellent (demonstrated real excellence and mastery in this area*). The quality ratings focus on therapists' demonstration of expertise, competence, and commitment; appropriate timing; clarity of language; and responding to where the client appears to be. For the purposes of the present study, only the quality ratings were used.

The Assessment subscale evaluates the extent to which a therapist assesses clients' alcohol and drug use, general level of functioning, current level of family or social support, and psychiatric symptoms. The General Support subscale measures therapist empathy, quality of the therapeutic relationship, and the degree to which the therapist provided support for the client. The Goals for Treatment subscale assesses the degree to which the therapist facilitated discussion of client goals for treatment. The Clinical Management subscale measures skills related to providing the "common factors" of psychotherapy and monitoring compliance with the study medications. The Twelve-Step Facilitation subscale measures the extent to which Twelve-Step participation was encouraged and includes interventions such as encouraging clients to attend self-help meetings, directly confronting client denial, and discussing the disease model of addiction. The Cognitive-Behavioral subscale measures the teaching of coping skills and monitoring and evaluating client thoughts.

Carroll et al. (2000) reported intraclass correlation coefficients (ICCs) with a sample of 19 sessions each rated by five raters. For adherence ratings, the ICCs ranged from .80 (Assessment) to .95 (Clinical Management), and for the competence ratings the ICC ranged from .71 (General Support) to .98 (Clinical Management), indicating acceptable to good reliability.

The construct validity of the YACS was examined by Carroll et al. (2000) through a confirmatory factor analysis to separately

evaluate the hypothesized factor structure for each subscale of the measure. The goodness-of-fit indices for the adherence subscales ranged from .92 on Clinical Management to .99 for Assessment. Carroll et al. (2000) also examined the convergent and discriminant validity by evaluating the relationship of the YACS subscales with four other therapy alliance measures and found that correlations with these various measures were generally in the expected directions.

Procedure

Each rater independently evaluated all 50 of the study audiotapes and reviewed only 2 tapes consecutively so as to avoid rater fatigue. One rater evaluated tapes using the MISTS, the second rater reviewed each tape using the YACS, and the third rater evaluated each tape using the YACS in addition to the MISTS.

Results

Means and standard deviations for the 16 MISTS items for each individual rater and the raters as a group are presented in Table 2. The reliability of the data derived with the MISTS was examined using generalizability theory. Generalizability theory, introduced by Cronbach and his colleagues (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Cronbach, Nageswari, & Gleser, 1963), provides a means for researchers to significantly improve their estimates of interrater reliability over other indices developed under classical test theory. Unlike other classical test theory coefficients, which index only a single source of error, generalizability coefficients, which are ICCs, allow test developers to represent and adjust for multiple sources of error in a single analysis and are easily estimated using variance components derived from analysis of variance design models. As with all ICCs, generalizability coefficients range from 0 to 1, with values nearer to 1 being most desirable. The resulting coefficient is interpreted as an index of the degree of association between the study's raters' ratings and the average ratings of the population of all possible raters. For this study, generalizability coefficients were estimated separately by item and across all items on the MISTS. As seen in Table 3, none of the MISTS items would be classified as poor according to Cicchetti's (1994) classification of clinical significance for interrater reliability (i.e.,

a coefficient less than .40). The generalizability coefficient between the two raters using the MISTS was high (ρ^2 = .79), which would be considered excellent. The generalizability coefficient for the two raters using the YACS was slightly higher (ρ^2 = .82), which would also be considered excellent. These results suggest that there was relatively strong agreement between raters using both measures.

Correlation coefficients were calculated to examine the convergent and discriminant validity between the total score on the MISTS and the six subscales of the YACS. We hypothesized that there would be a positive relationship between the MISTS total score and the YACS Assessment, Support, and Goals subscales. We also hypothesized that there would be no relationship between the MISTS total score and the Clinical Management, Twelve-Step Facilitation, and Cognitive-Behavioral Management subscales of the YACS. As seen in Table 4, positive correlations were found between the MISTS total score and the Support and Goals subscales of the YACS, which supports the hypotheses regarding these measures. A relatively weak correlation between the Assessment subscale and the MISTS total score did not support the hypothesized convergent validity of these scales. A weak negative correlation was found between the MISTS total score and the Twelve-Step Facilitation subscale of the YACS. A weak positive correlation was found between the MISTS total score and the Clinical Management scale. These two findings lend support for the a priori hypotheses. However, a moderate positive correlation was found between the MISTS total score and the Cognitive-Behavioral subscale of the YACS, which was in contrast to the hypotheses regarding these measures.

Discussion

Instruments that assess both adherence to particular forms of ESTs as well as the quality with which therapists implement those therapies are needed in order for psychotherapy research and training to advance. Instruments that emphasize only adherence to a treatment modality miss a critical aspect of the delivery of mental health services, which is highly relevant for research as well as clinical training and supervision purposes. Therefore, we developed an instrument that measures both adherence to, and quality of

implementation for, MI therapy, an important treatment modality in the addictions field as well as a variety of other areas. We also aimed to develop an instrument with high clinical utility as well as acceptable support regarding its psychometric properties.

The results of this study suggest that the data obtained with the MISTS are reliable and reasonably valid. Acceptable interrater consistency was found between raters who independently used the MISTS to analyze audiotaped therapy sessions in which therapists incorporated MI. This suggests that the MISTS has the potential to provide reliable data when used in similar ways to the way it was used in this study.

The results across the individual items on the MISTS showed significant variability, however. In particular, there were five items (simple reflection, complex reflection, engaging the client in treatment, addressing client ambivalence, and rolling with resistance) that produced lower ICCs, ranging from .41 to .55. Although these coefficients are considered "fair" according to Cicchetti's (1994) classification of clinical significance, clearly the instrument would benefit if they were stronger. Several reasons may account for the lower consistency on these items. Motivational interviewing is a complex treatment approach in which there is continued uncertainty regarding the "active ingredients" in the intervention (Longabaugh, 2001; Rollnick, 2001). Several of the items that produced less consistent results in this study were directly related to the theorized ingredients of MI but involve fairly complex constructs (e.g., complex reflection, addressing client ambivalence, and rolling with resistance). Even though steps to ensure content validity were implemented, the fact that some of these concepts are relatively complex led to complex behavioral anchors for some items, which were multidimensional, thus reducing the content validity for these items. Similarly, the fact that MI is described as a treatment style as opposed to a specific set of skills may make some of these concepts more difficult to define operationally and thus monitor. For example, the midpoint anchor for the rolling-with-resistance item is "acknowledges resistance, argues minimally with client, but seems to lack skill in shifting focus during session." It is also possible that the training provided to the raters in this study was too brief, or that raters more experienced with MI would provide more reliable data. Although the reliability coefficients

obtained for each of the items in considered acceptable from the perspective of clinical significance, more investigation of these issues is warranted.

An additional concern in this study is potential rater bias, as one rater consistently rated sessions lower on 15 out of 16 MISTS items. However, the amount of disagreement between the raters was not that large (often just one tenth of a standard deviation). The rater bias in this case can only be construed as a tendency for the raters to use the scales in a consistent but idiosyncratic fashion, with one tending to score higher and the other lower. Neither one can be considered to be wrong outside of some absolute standard, which we are lacking. This is an unfortunate but typical occurrence when dealing with raters. On the other hand, the whole point of the generalizability coefficient is to index the relationship of our raters to the average of the hypothetical population of all possible raters, allowing for inferences to be made to other raters despite such differences. In the formation of this index we have adjusted for such differences in the rater's ratings. However the fact that one rater consistently rated higher or lower as the case may be is problematic in that ideally there would be better consistency between raters. Future development of the MISTS will need to include stronger, more well-defined behavioral anchors and more extensive training of raters to improve agreement between raters.

The study results also provide evidence supporting the convergent and discriminant validity of the MISTS. As hypothesized, there was a strong positive correlation between the MISTS total scale and the General Support and Goals for Treatment subscales of the YACS, an important finding given that three important components of MI are (a) supporting the client, (b) helping the client resolve ambivalence, and (c) helping the client establish goals in relation to a problem behavior (Miller & Rollnick, 2002). In contrast, a weak positive correlation between the MISTS and the Assessment subscale of the YACS did not support the hypothesis regarding convergent validity. Examining the responses of the Assessment subscale items revealed little variability across the items, which is likely the result of the sample sessions being primarily therapy focused with little formal assessment. As a result, the Assessment subscale items had only weak correlations with the MISTS total score. Future investigation of the convergent and divergent validity of the MISTS should focus more on

the particular types of sessions being rated. The possibility that the MISTS might be strengthened through the addition of assessment-related items should also be investigated.

Also contrary to what was hypothesized, we found a statistically significant positive correlation with a small to medium effect size (r = .53, r^2 = .28) between the MISTS and the Cognitive-Behavioral Management subscale of the YACS. This is somewhat problematic as we had hypothesized that there would be no correlation between the MISTS and this subscale because there are important differences between these two interventions. Specifically, cognitive-behavioral therapy is intended to be more directive, and the therapist takes on more of an expert role, does more thorough assessment of and education regarding substance use behavior, identifies and challenges faulty cognitions, and teaches coping skills that can be used to substitute for substance use behavior (Carroll, 1999). Although the correlation between the MISTS and the Cognitive-Behavioral Management subscale is substantial, it is important to note that the two scales share only 28% of the variance. A post hoc analysis of the correlations between the items on the two scales found that a high degree of the correlation was largely due to one item in the Cognitive-Behavioral Management subscale (i.e., "Discussing any high risk situations the patient encountered in the past and exploring specific actions taken to avoid or cope with the situation(s)"; r = .55, p = .01) and, to a lesser extent, a second item in that subscale (i.e., "Exploring specific cravings, triggers, or urges for use"; r = .32, p = .05). The correlation of these items with the MISTS would not be unexpected given that an important component of MI involves exploring previous drinking situations and the positive or negative experiences associated with those situations (Miller & Rollnick, 2002). Consequently, the moderate level of correlation found between these two measures is of somewhat less concern than it first appears.

The use of the MISTS in MI intervention research can help address a significant threat to the internal validity of a study by providing an assessment of treatment fidelity and quality (Calsyn, 2000; Waltz, Addis, Koerner, & Jacobson, 1993). This will address the problem of treatment monitoring identified by Miller (2001) and Rollnick (2001) by ensuring that MI is being implemented as it was designed to be practiced. The MISTS can also be used in answering the

call for process research in relation to establishing ESTs (Wampold, Lichtenberg, & Waehler, 2002). For example, the MISTS can be used with other measures (e.g., stage of change, client satisfaction, treatment outcome) to help identify the process of client change when using MI and determine which skills consistent with the spirit of MI may facilitate client change better than others.

Data obtained with the MISTS may also be quite useful in the training and supervision of therapists using MI, and the emphasis on clinical utility in the design of the instrument may increase the likelihood that it will actually be used for this purpose. The raters in this study considered the behavioral count section easy to complete in one continuous viewing of a tape. These frequency counts of important MI therapist responses, along with the global ratings of the quality of a therapist's intervention, provide supervisors with data for addressing specific skills and more global characteristics critical to the successful implementation of MI. Asking therapist trainees to review and rate their own sessions using the MISTS may also help them develop better MI skills. Having student therapists review their own taped therapy sessions has been found to be quite helpful when they are learning core counseling skills, and this practice is likely to prove helpful when learning MI as well (Bernard & Goodyear, 1998).

Although the MISTS may provide a useful instrument for addressing a number of research, training, and supervision needs with regard to the use of MI, the reliability and validity of the instrument should receive more thorough evaluation. The instrument should be cross-validated with multiple samples and in multiple settings, and with special attention paid to the items with lower reliability estimates. Given that the therapeutic alliance is an important aspect of MI, more thorough examination of the validity of the MISTS could come from studies comparing the MISTS with other measures of the therapeutic alliance to examine questions regarding convergent validity. The MISTS should be compared with measures assessing different therapy approaches to examine discriminant validity. In addition, the MISC has recently been revised in an attempt to address some of the concerns previously mentioned (Miller, Moyers, Ernst, & Amrhein, 2003), although we were unable to find any psychometric evaluation of this measure. Directly comparing this alternative measure of the same

constructs will also provide important data for addressing the strengths and weaknesses of the MISTS.

References

Baer, J. S., Kivlahan, D. R., Blume, A. W., McKnight, P., & Marlatt, G. A. (2001). Brief intervention for heavy-drinking college students: 4-year follow-up and natural history. *American Journal of Public Health*, *91*, 1310–1316.

Barrett, D., Rugg, C., Zweben, A., Campbell, T. C., & Madson, M. B. (1998). *Motivational enhancement therapy plus network support and skills enhancement: A manual for alcohol outpatient treatment*.Unpublished treatment manual.

Barsky, A., & Coleman, H. (2001). Evaluating skill acquisition in motivational interviewing: The development of an instrument to measure practice skills. *Journal of Drug Education*, *31*, 69–82.

Bein, T. H., Miller, W. R., & Boroughs, J. M. (1993). Motivational interviewing with alcohol outpatients. *Behavioural and Cognitive Psychotherapy*, *21*, 347–356.

Berg-Smith, S. M., Stevens, V. J., Brown, K. M., Van Horn, L., Gernhofer, N., Peters, E., et al. (1999). A brief motivational intervention to improve dietary adherence in adolescents. *Health Education Research: Theory and Practice*, *14*, 399–410.

Bernard, J. M., & Goodyear, R. K. (1998). *Fundamentals of clinical supervision* (2nd ed.). Boston: Allyn & Bacon.

Brown, J. M., & Miller, W. R. (1993). Impact of motivational interviewing on participation and outcome in residential treatment. *Psychology of Addictive Behaviors*, *7*, 211–218.

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology*, *71*, 843–861.

Butler, C. C., Rollnick, S., Cohen, D., Bachmann, M., Russell, I., & Stott, N. (1999). Motivational consulting versus brief advice for smokers in general practice: A randomized trial. *British Journal of General Practice*, *49*, 611–616.

Calsyn, R. J. (2000). A checklist for critiquing treatment fidelity studies. *Mental Health Services Research*, *2*, 107–113.

Carey, M. P., Kalichman, S. C., Forsyth, A. D., Wright, E. M., & Johnson, B. T. (1997). Enhancing motivation to reduce the risk of HIV infection for economically disadvantaged urban women. *Journal of Consulting and Clinical Psychology*, *65*, 531–541.

Carroll, K. M. (1999). Behavioral and cognitive behavioral treatments. In B. S.McCrady & E. E.Epstein (Eds.), *Addictions: A comprehensive guidebook* (pp. 250–267). New York: Oxford University Press.

Carroll, K. M., Farentinos, C., Ball., S. A., Crits-Christoph, P., Libby, B., Morgernstern, J., et al. (2002). MET meets the real world: Design issues and clinical strategies in the clinical trials network. *Journal of Substance Abuse Treatment*, *23*, 73–80.

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*, 225–238.

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

Cigrang, J. A., Severson, H. H., & Peterson, A. L. (2002). Pilot evaluation of a population-based health intervention for reducing use of smokeless tobacco. *Nicotine and Tobacco Research*, *4*, 127–131.

Colby, S. M., Monti, P. M., Barnett, N., Rohsenow, D. J., Weissman, K., Spirito, A., et al. (1998). Brief motivational interviewing in a hospital setting for adolescent smoking: A preliminary study. *Journal of Consulting and Clinical Psychology*, *3*, 574–578.

Connors, G. J., Walitzer, K. S., & Dermen, K. H. (2002). Preparing clients for alcoholism treatment: Effects on treatment participation and outcomes. *Journal of Consulting and Clinical Psychology*, *70*, 1161–1169.

Corvino, J., Carroll, K., Nuro, K., Nich, C., Sifry, R., Frankforter, T., et al. (2000). *Yale Adherence and Competence Scale guidelines*. West Haven, CT: Yale University Psychotherapy Development Center.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cronbach, L. J., Nageswari, R., & Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *British Journal of Statistical Psychology*, *16*, 137–163.

Davison, G. C. (1998). Being bolder then the Boulder model: The challenge of education and training in empirically supported treatments. *Journal of Consulting and Clinical Psychology*, *66*, 163–167.

Dunn, C., Deroo, L., & Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review: Reply. *Addiction*, *96*, 1774–1775.

Emmons, K. M., Hammond, S. K., Fava, J. L., Velicer, W. F., Evans, J. L., & Monroe, A. D. (2001). A randomized trial to reduce passive smoke exposure in low-income households with young children. *Pediatrics*, *108*, 18–24.

Handmaker, N. S., Miller, W. R., & Manicke, M. (1999). Findings of a pilot study of motivational interviewing with pregnant drinkers. *Journal of Studies on Alcohol*, *60*, 285–288.

Hill, C. E., & O'Brien, K. M. (1999). *Helping skills: Facilitating exploration, insight, and action*. Washington, DC: American Psychological Association.

Ivey, A. E., & Bradford Ivey, M. (2003). *Intentional interviewing and counseling: Facilitating client development in a multicultural society*. Pacific Grove, CA: Brooks/Cole.

Lincourt, P., Kuettel, T. J., & Bombardier, C. H. (2002). Motivational interviewing in a group setting with mandated clients: A pilot study. *Addictive Behaviors*, *27*, 381–391.

Longabaugh, R. (2001). Why is motivational interviewing effective?*Addiction*, *96*, 1773–1774.

Marlatt, G. A., Baer, J. S., Kivlahan, D. R., Dimeff, L. A., Larimer, M. E., Quigley, L. A., et al. (1998). Screening and brief intervention for high-risk college student drinkers: Results from a 2-year follow-up

assessment. *Journal of Consulting and Clinical Psychology*, *66*, 604–615.

Martino, S., Carroll, K. M., O'Malley, S. S., & Rounsaville, B. J. (2000). Motivational interviewing with psychiatric ill substance abusing patients. *American Journal on Addictions*, *9*, 88–91.

Miller, W. R. (2000). *Motivational interviewing skill code (MISC): Coder's manual*. Unpublished manual, University of New Mexico. Available at: http://www.motivationalinterview.org/

Miller, W. R. (2001). When is it motivational interviewing?*Addiction*, *96*, 1770–1771.

Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior?*Behavioral and Cognitive Psychotherapy*, *29*, 457–471.

Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2003). *Manual for the motivational interviewing skill code* (2nd ed.). Unpublished manual, University of New Mexico. Available at: http://www.motivationalinterview.org/

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people to change* (2nd ed.). New York: Guilford Press.

Miller, W. R., Sovereign, R. G., & Krege, B. (1988). Motivational interviewing with problem drinkers: II. The drinker's check-up as a prevention intervention. *Behavioural Psychotherapy*, *16*, 251–268.

Miller, W. R., Yahne, C. E., & Tonigan, J. S. (2003). Motivational interviewing in drug abuse services: A randomized trial. *Journal of Consulting and Clinical Psychology*, *71*, 754–763.

Miller, W. R., Zweben, A., DiClemente, C. C., & Rychtarik, R. G. (1995). *Motivational enhancement therapy manual: A clinical research guide for therapists treating individuals with alcohol abuse and dependence* (Project MATCH Monograph Series, *Vol. 2*, NIH Publication No. 94-3723).Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Monti, P. M., Colby, S. M., Barnett, N. P., Spirito, A., Rohsenow, D. J., Meyers, M., et al. (1999). Brief intervention for harm reduction with

alcohol-positive adolescents in a hospital emergency room. *Journal of Consulting and Clinical Psychology*, *67*, 989–994.

Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skill code. *Behavioural and Cognitive Psychotherapy*, *31*, 177–184.

Picciano, J. F., Roffman, R. A., Kalichman, S. C., Rutledge, S. E., & Berghuis, J. P. (2001). A telephone based brief intervention using motivational enhancement to facilitate HIV risk reduction among MSM: A pilot study. *AIDS and Behavior*, *5*, 251–262.

Rollnick, S. (2001). Enthusiasm, quick fixes, and premature controlled trials. *Addiction*, *96*, 1769–1770.

Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing?*Behavioural and Cognitive Psychotherapy*, *23*, 325–334.

Schneider, R. J., Casey, J. C., & Kohn, R. (2000). Motivational versus confrontational interviewing: A comparison of substance abuse assessment practices at employee assistance programs. *Journal of Behavioral and Health Services and Research*, *27*, 60–74.

Stein, M. D., Charuvastra, A., Maksad, J., & Anderson, B. J. (2002). A randomized trial of a brief alcohol intervention for needle exchangers (BRAINE). *Addiction*, *97*, 691–700.

Stotts, A. L., DiClemente, C. D., & Dolan-Mullen, P. (2002). One-to-one: A motivational intervention for resistant pregnant smokers. *Addictive Behaviors*, *27*, 275–292.

Stotts, A. L., Schmitz, J. M., Rhoades, H. M., & Grabowski, J. (2001). Motivational interviewing with cocaine-dependent patients: A pilot study. *Journal of Consulting and Clinical Psychology*, *69*, 858–862.

Substance Abuse and Mental Health Services Administration. (1999). *Enhancing motivation for change in substance abuse treatment*. (DHHS Publication No. SMA 99-3354). Rockville MD: CDM Group.

Tappin, D. M., McKay, C., McIntyre, D., Gilmour, W. H., Cowan, S., Crawford, F., et al. (2000). A practical instrument to document the process of

motivational interviewing. *Behavioral and Cognitive Psychotherapy*, *28*, 17–32.

Waehler, C. A., Kalodner, C. R., Wampold, B. E., & Lichtenberg, J. W. (2000). Empirically supported treatments (ESTs) in perspective: Implications for counseling psychology training. *The Counseling Psychologist*, *28*, 657–671.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620–630.

Wampold, B. E., Lichtenberg, J. W., & Waehler, C. A. (2002). Principles of empirically supported interventions in counseling psychology. *The Counseling Psychologist*, *30*, 197–217.

Wertz, J. S. (1994). The effect of motivational interviewing on treatment participation, self-efficacy and alcohol use at follow-up in inpatient alcohol dependent adults. (Doctoral dissertation, VA Polytechnic Institute and State University, 1994). *Dissertation Abstracts International*. p. 55.1B

Woollard, J., Beilin, L., Lord, T., Puddey, I., MacAdam, D., & Rouse, I. (1995). A controlled trial of nurse counseling on lifestyle change for hypertensives treated in general practice: Preliminary results. *Clinical and Experimental Pharmacology and Physiology*, *22*, 466–468.

Yahane, C. E., Miller, W. R., Irvin-Vitela, L., & Tonigan, J. S. (2002). Magdalena pilot project: Motivational outreach to substance abusing women sex workers. *Journal of Substance Abuse Treatment*, *23*, 49–53.

Appendix

Table 1
Motivational Interviewing Studies Reviewed With Regard to Adherence to, Training in, and Monitoring of the Intervention

Study	None	Described training and supervision	Intervention training provided, no supervision	Intervention training, supervision tool—no psychometrics	Intervention training, supervision tool—with psychometrics
Alcohol and other drugs					
Miller et al. (1988)	X				
Brown & Miller (1993)	X				
Bein et al. (1993)	X				
Wertz (1994)			X		
Handmaker et al. (1999)	X				
Marlatt et al. (1998)		X			
Monti et al. (1999)				X	
Schneider et al. (2000)				X	
Martino et al. (2000)	X				
Stotts et al. (2001)				X	
Baer et al. (2001)		X			
Connors et al. (2002)					X
Stein et al. (2002)				X	
Lincourt et al. (2002)	X				
Miller, Yahne, & Toningan (2003)					X
Tobacco					
Colby et al. (1998)	X				
Butler et al. (1999)	X				
Emmons et al. (2001)		X			
Cigrang et al. (2002)	X				
Stotts et al. (2002)		X			
Medical/health					
Woollard et al. (1995)	X				
Berg-Smith et al. (1999)		X			
Carey et al. (1997)	X				
Belcher et al. (1998)	X				
Picciano et al. (2001)	X				
Yahane et al. (2002)	X				
Total	14 (54%)	5 (19%)	1 (4%)	4 (15%)	2 (8%)

Table 2
Means and Standard Deviations for Each Item on the Motivational Interviewing Supervision and Training Scale

Item	Rater (s)	*M*	*SD*
Questions	1 and 3	4.44	1.56
	1	4.36	1.30
	3	4.52	1.79
Simple reflection	1 and 3	4.32	1.76
	1	3.98	1.63
	3	4.66	1.84
Complex reflection	1 and 3	2.69	1.83
	1	2.40	1.62
	3	2.98	1.98
Affirming	1 and 3	4.48	1.96
	1	4.06	2.01
	3	4.90	1.84
Summarization	1 and 3	3.60	1.98
	1	3.60	1.80
	3	3.60	2.15
Engaging client	1 and 3	5.28	1.60
	1	5.02	1.58
	3	5.54	1.59
Elicits/reinforces change talk	1 and 3	4.32	1.85
	1	3.94	1.85
	3	4.70	1.79
Addresses ambivalence	1 and 3	4.22	1.56
	1	3.44	1.51
	3	5.00	1.17
Rolls with resistance	1 and 3	4.33	1.67
	1	3.54	1.55
	3	5.12	1.40
Collaborating with client	1 and 3	4.48	1.70
	1	4.16	1.77
	3	4.80	1.57
Supports self-efficacy	1 and 3	4.92	1.74
	1	4.50	1.89
	3	5.34	1.47
Use of active listening skills	1 and 3	4.23	1.66
	1	4.08	1.39
	3	4.38	1.89
Sequence of motivational interview	1 and 3	3.65	1.73
	1	3.22	1.62
	3	4.08	1.74
Spirit of motivational interview	1 and 3	4.71	1.79
	1	4.20	1.84
	3	5.22	1.60
Response of client	1 and 3	5.19	1.55
	1	4.92	1.44
	3	5.46	1.63
Effectiveness of therapist	1 and 3	4.27	1.86
	1	3.96	1.66
	3	4.58	2.01

Note. Item responses can range from 1 (*poor use of skills*) to 7 (*excellent use of skills*). The *n* for individual raters is 50, and the *n* for Raters 1 and 3 combined is 100.

Means and Standard Deviations for Each Item on the Motivational Interviewing Supervision and Training Scale

Table 3

Intraclass Correlations (ICCs) for Individual Items on the Motivational Interviewing Supervision and Training Scale and Classification of Clinical Significance

Item	ICC	Classification[a]
Questions	.71	good
Simple reflection	.55	fair
Complex reflection	.41	fair
Affirming	.81	excellent
Summarization	.73	good
Engaging client	.46	fair
Elicits/reinforces change talk	.65	good
Addresses ambivalence	.45	fair
Rolls with resistance	.55	fair
Collaborating with client	.68	good
Supports self-efficacy	.74	good
Use of active listening skills	.71	good
Sequence of motivational interview	.71	good
Spirit of motivational interview	.66	good
Response of client	.70	good
Effectiveness of therapist	.76	excellent

[a] According to Cicchetti (1994).

Table 4

Correlation Between Motivational Interviewing Supervision and Training Scale Total and Yale Adherence and Competence Scale Subscales

Subscale	r	r^2
Convergent		
Assess	.29*	.08
Support	.72**	.52
Goals	.70**	.49
Discriminant		
Clinical Management	.15	.02
Twelve-Step Facilitation	−.27	.07
Cognitive–Behavioral Management	.53**	.28

Note. $N = 100$.
* $p = .05$. ** $p = .01$.