Regular article

Measures of fidelity in motivational enhancement: A systematic review

Michael B. Madson, (Ph.D.)[a],[*], Todd C. Campbell, (Ph.D.)[b]

[a]Department of Psychology, The University of Southern Mississippi, Hattiesburg, MS, USA
[b]The Center for Addiction and Behavioral Health Research, Marquette University, Milwaukee, WI, USA

Abstract

The movement to use empirically supported treatments has increased the need for researchers and clinical supervisors to evaluate therapists' adherence to and competence in particular empirically supported interventions. Motivational interviewing (MI) is an empirically supported intervention for substance abuse and other behavioral problems. However, for this intervention to work, it must be provided with fidelity and skill. This article provides a systematic review of MI adherence and competence measures that have been developed and described independently elsewhere. Recommendations for refinement of the measures and future research are also discussed. © 2006 Elsevier Inc. All rights reserved.

Keywords: Treatment fidelity; Motivational interviewing; Training and supervision; Substance abuse treatment

1. Introduction

The demand for evidence-based practice in the substance use disorder field has required clinicians, researchers, and educators to focus on how knowledge is best transferred from research to professional practice (Campbell, Catlin, & Melchert, 2003; Campbell, Daood, Catlin, & Moskovitz, 2005; Institute of Medicine, 1998). As a result, the field is advancing in addressing the issue of adherence and competence regarding how interventions are being implemented. For example, Carroll et al. (2002) suggested that a major challenge faced in technology transfer involves ensuring therapists' adherence to and competence in the implementation of interventions through assessment. This form of assessment will be particularly important in examining the change process to help understand how interventions work, find ways to improve them, and learn how to overcome barriers to implementation and enhance future execution of the intervention as these are vital factors emerging in relation to evidence-based practice (Orwin, 2000).

Evidence-based practice also calls for educators and clinical supervisors to provide more rigorous evaluations of clinicians that include adherence and competence assessment (Bernard & Goodyear, 1998). Furthermore, Falender and Shafranske (2004) suggested that assessment of competent practice would be enhanced by establishing assessment tools that are operationalized to yield data with good reliability and validity. Quality measurement tools would assist many supervisors in conducting empirically based evaluations, as opposed to relying solely on clinical experience. These instruments are especially important when an intervention involves complex concepts. Motivational interviewing (MI), for example, which is a complex behavioral intervention often used with substance use disorders, would benefit from empirically based evaluation tools.

Motivational interviewing is a directive client-centered approach for eliciting behavior change by assisting clients in exploring and resolving ambivalence (Miller & Rollnick,

* Corresponding author. Department of Psychology, The University of Southern Mississippi, 118 College Drive No. 5025, Hattiesburg, MS 39406-5025, USA.

E-mail address: michael.madson@usm.edu (M.B. Madson).

2002) that shows promise as an efficacious intervention with a variety of behaviors in various treatment contexts (Burke, Arkowitz, & Menchola, 2003). However, an increasing concern reported in several studies on MI is the fidelity and quality with which an intervention was implemented. For example, Madson, Campbell, Barrett, Brondino, and Melchert (2005) found variability with which studies examining MI described the training, supervision, and monitoring of therapists. This finding 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 MI fidelity through the direct monitoring of the intervention as opposed to clinician self-report (see also Carroll et al., 2002).

Several measures have been developed to address the increasing demand for evaluation of therapists implementing MI. The purpose of this study was to expand on previous work that we have done in this area by providing a comprehensive and current review of MI adherence and competence measures that have been developed and described independently elsewhere. There has previously been no consolidated evaluation of these measures. In addition, we build on previous work by providing suggestions for future directions in developing MI adherence and competence measures, with particular attention on how revisions can make them more amenable in supervision and training settings. This review summarizes what has been accomplished thus far in the development of these measures and provides information that can assist researchers, educators, and clinical supervisors in developing and selecting empirically sound tools to conduct these needed evaluations. It provides a description of each measure, reliability and validity estimates, and evaluation.

2. Materials and Methods

To identify measures for inclusion in this review, we conducted a literature search using the PsychINFO, Health and Psychosocial Instruments, and Medline databases. Database terms used included "motivational interviewing," "motivational enhancement therapy," "therapist fidelity measures," "adherence and competence measures," "therapist competence," and "technology transfer." This method provided a variety of measures. Measures were included in this review if they directly stated that they were developed for MI or upon review measured, indirectly, skills important to MI. This process resulted in five measures for inclusion in this review: the Yale Adherence and Competence Scale (YACS); the Motivational Interviewing Skill Code (MISC); the Motivational Interviewing Process Code (MIPC); the Motivational Interviewing Treatment Integrity Scale (MITI); and the Motivational Interviewing Supervision and Training Scale (MISTS).

2.1. Yale Adherence and Competence Scale

2.1.1. Measure description

The YACS is a 50-item measure designed to evaluate clinician adherence to and competence in implementing interventions common among most therapies as well as interventions associated with specific therapy modalities (Corvino et al., 2000). The instrument includes three subscales (assessment, general support, and goals for treatment) to assess general interventions common across therapies and three subscales (clinical management, 12-step facilitation, and cognitive–behavioral management) to 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*); quality ratings, from 1 (*very poor* = therapist handled this in an unacceptable, even toxic, manner) to 7 (*excellent* = therapist demonstrated real excellence and mastery in this area).

2.1.2. Psychometric properties

Carroll et al. (2000) reported intraclass correlation coefficients (ICCs) that ranged from .80 to .95 for adherence and from .71 to .98 for competence. Madson et al. (2005) also found that the YACS quality ratings can be used reliably ($\rho^2 = .82$). The construct validity of the YACS was examined by Carroll et al. through a confirmatory factor analysis, and the goodness-of-fit indices for the adherence subscales ranged from .92 to .99. Furthermore, Madson et al. found evidence for convergent and discriminant validity when the YACS was compared with the MISTS (see below).

2.1.3. Evaluation

The YACS can be used consistently among raters to assess both general and specific psychotherapeutic interventions. It is apparent that Corvino et al. (2000) made great efforts to develop a measure that can be used to assess clinician adherence to and quality in implementing several psychotherapies. The YACS has specific utility for use in research studies in which multiple interventions are being implemented.

Although the original YACS, in which the psychometric data cited apply, assessed elements common to all psychotherapies, it did not directly assess the principles that are essential to the spirit of MI such as eliciting/reinforcing change talk and rolling with resistance. However, a recently revised version includes nine items that are intended to assess MI concepts such as rolling with resistance and MI spirit, although no psychometric information was found.

Currently, the YACS is intended as a research tool and it may behoove researchers to consider how this measure may be modified to enhance its utility in training and supervision contexts. For example, general behavioral anchors were developed to be used across all items. Developing specific behavioral anchors for each item may improve the utility of the measure in training and supervision settings by

enhancing the feedback that supervisors could provide to clinicians through comparison with concrete behaviors.

2.2. Motivational Interviewing Skill Code

2.2.1. Measure description

The MISC was developed for use with audiotape or videotape MI sessions to encode interactions between therapists and clients (Miller, 2000). It requires the rater to conduct three rating passes. The first pass involves the rater reviewing the entire interview and completing the Global Therapist Rating (acceptance, egalitarianism, empathy, genuineness, warmth, and spirit of MI), the Global Client Rating (affect, cooperation, disclosure, and engagement), and the Global Interaction Rating (level of collaboration and benefit of the interaction) scales using a 7-point Likert-type scale after the review. The second pass requires the rater to classify every therapist and client utterance into a specific mutually exclusive category. For example, therapist utterances can be classified as affirming, confronting, questioning (closed or open), reflecting (repeat, rephrase, paraphrase, summarize), or reframing; client utterances, as asking questions, neutrally responding, resisting change, or change talking. Finally, during the third pass, the rater computes the talk time during the session for both the therapist and the client.

2.2.2. Psychometric properties

Tappin et al. (2000) estimated ICCs as .39 for the therapist scale, .53 for the client scale, and .51 for the interaction scale. Moyers et al. (2003) found these ICCs to range from .25 to .79 for the MISC global items and from .00 to 1.00 for the behavioral counts. Miller and Mount (2001) examined the construct validity of the MISC by having independent raters rate MI sessions conducted by four experts in MI. The expectation was that the experts in MI would rate highly on the MISC. This was indeed the case as the expected responses were consistently recorded by the independent raters; however, interrater correlations were not reported. Furthermore, deJonge, Schippers, and Schaap (2005) found that the general principles of MI are covered by the MISC, in an unbalanced fashion, however.

2.2.3. Evaluation

The MISC has potential to be a measure that can be used consistently by raters and provides information that it was intended to provide. Specifically, the MISC appears useful for deconstructing the interaction between clients and therapists. However, further examination of the psychometric properties is needed, specifically its construct validity. Nevertheless, the MISC demonstrates some promise as a tool that can be used in research to quantify a therapist's adherence to MI. Although adherence to the tenets of the intervention is an important component in learning and implementing it, the competence with which the intervention is implemented is also an important component in

training therapists. Therefore, any measure to be used as a training and supervision aid would benefit from evaluating the quality with which the intervention was implemented. Similarly, although the MISC shows promise as a measure to be used in a research setting, its complexity makes it more difficult to be implemented in a training or practice setting—and many supervisors may prefer a briefer and less complex measure. For example, Tappin et al. (2000) suggested that it could take up to 4 hours to evaluate a single session using the MISC, which is time that educators and practitioners may not be able to commit.

Miller, Moyers, Ernst, and Amrhein (2003) recently developed a revised version, the MISC 2.0. This updated version no longer requires the third pass for calculation of clinician and client talk time. At this time, there is no information available on the reliability and validity of the MISC 2.0.

2.3. Motivational Interviewing Process Code

2.3.1. Measure description

The MIPC was designed to provide a measure of the primary skills consistent with and contrary to MI to aid the teaching and evaluation of learning MI skills (Barsky & Coleman, 2001). It consists of two subscales: The first subscale, functional skills, includes 13 items that are rated on a Likert-type scale ranging from 1 (*component not demonstrated*) to 5 (*outstanding*). Items range from basic microcounseling skills such as expressing empathy to skills associated with the spirit of MI such as amplifying discrepancies and ambiguity. The second subscale, dysfunctional skills, includes 12 items also rated on a 5-point Likert-type scale that ranges from 1 (*demonstrates throughout the interview*) to 5 (*avoids completely*). Items include MI-inconsistent behaviors such as arguing/debating with the client and labeling the client. The authors used a modified Delphi approach in which practitioners with strong experience in MI completed a questionnaire asking them to identify attributes required for effective MI and behavior considered counter to MI (Barsky & Coleman, 2001). Next, the authors conducted three focus groups in which participants reviewed the results of the survey and devised a consolidated list of MI positive and dysfunctional skills. This process resulted in the MIPC.

2.3.2. Psychometric properties

Barsky and Coleman (2001) estimated the interrater reliability through an analysis of the percentage of agreement between raters for the functional skills subscale (51.27%) and the dysfunctional skills subscale (75.03%). They suggested that the consensus that was achieved during the focus groups of professionals is evidence for construct validity. They also suggested that consistency between the language used by practitioners in the focus groups and that used in the literature is further evidence of validity.

2.3.3. Evaluation

The MIPC represents a good initial attempt to develop a tool for use in training to assess learning of MI skills. Barsky and Coleman (2001) described a comprehensive process used in developing items for the instrument that included surveying professionals identified as experts in MI and conducting focus groups. Furthermore, the authors included skills that were considered inconsistent with the MI framework, which is important in relation to training therapists. Specifically, inclusion of skills inconsistent with MI will help supervisors and therapists point out skills that need to be modified or changed to comply better with the spirit of MI.

Although the MIPC in its current form represents a first attempt at developing a measure for training clinicians in MI, several limitations must be addressed before use in education and training settings. Most notable of these limitations is the method used to estimate the interrater reliability. Although, as the authors stated, calculating the percentage of agreement between raters is a method often used to estimate interrater reliability, limitations of this method must be considered (Cronbach, Nageswari, & Gleser, 1963). Specifically, percentage of agreement does not account for agreement caused by chance and, as a result, can inflate the estimate of reliability.

Another limitation that must be considered relates to the professionals surveyed as experts in MI. The authors did not define the criteria used to determine how one was considered an expert in MI. This fact may be problematic as the authors indicated that they chose to use language identified by the professionals rather than the language in the literature to operationalize the measure and we are unsure of how experienced the professionals were in MI. Both Miller (2001) and Rollnick (2001) expressed concern that some may misinterpret or not accurately understand the concepts or complexity of MI. As a result, the authors' conclusions about the validity of the measure would benefit from having the determination of expert status explicated and from including statistical methods for estimating validity, such as convergent and discriminant methods.

2.4. Motivational Interviewing Treatment Integrity Scale

2.4.1. Measure description

In an attempt to reduce the complexity in assessing MI with the MISC, Moyers, Martin, Manuel, Hendrickson, and Miller (2005) developed the MITI. The MITI consists of two components: global ratings and behavioral counts. Global ratings of the overall rater's judgment are measured on a 7-point Likert-type scale ranging from 1 (*low*) to 7 (*high*) for two items, empathy/understanding and spirit of MI, which are important MI components (Moyers et al., 2005). Behavioral counts are simple tallies of interviewer behaviors. These behaviors are giving information, MI adherent, MI nonadherent, questions (closed and open), and reflection (simple and complex). Moyers et al. identified the

six MITI items through an exploratory factor analysis of the MISC 1.0 to identify the underlying factors. This process resulted in the current version of the MITI.

2.4.2. Psychometric properties

Moyers et al. (2005) calculated ICCs to estimate the interrater reliability of the global ratings and found an ICC of .51 for empathy/understanding and that of .58 for spirit of MI. The ICCs for the behavioral counts ranged from .57 to .96. The validity of the MITI was calculated through a canonical correlation between the MISC exploratory factor analysis and the MITI items as the same tapes were reviewed using both measures. The authors concluded that these results suggest convergence between the MITI and the MISC.

2.4.3. Evaluation

The MITI shows promise in being used reliably to observe psychotherapy sessions using MI. Moyers et al. (2005) made a concerted effort to retain the core factors of the MISC while reducing the complexity and length of time associated with it. Furthermore, it appears as if the MITI, because of its reduced length and complexity as compared with the MISC, has promise for use in training and supervision settings in addition to research settings.

Although the MITI shows promise in rating MI sessions, some of its limitations need to be considered before it is used in education and training settings. It is important to recognize, as the authors noted, that the MITI appears to assess MI-relevant clinician attributes such as empathy and use of microskills but does not capture the intentional and strategic use of MI. This is an important limitation of the measure because it may not provide an adequate assessment of very important principles of MI such as eliciting change talk and can be problematic when one is trying to capture the overall therapist competence and fidelity in using the more complex MI skills. Furthermore, the reliability coefficients for the two global items, empathy/understanding and spirit of MI, are considered fair according to the classification of clinical significance by Cicchetti (1994). However, refinements could be made to strengthen these two items. For example, specific behavioral anchors could be used for different points on the Likert scale to help the rater in decision making. Making modifications such as the one proposed may help the MITI become a stronger measure by providing concrete information to aid raters. Furthermore, this modification would provide trainers and supervisors with descriptions of behavior in which to compare therapists' behavior to foster learning.

2.5. Motivational Interviewing Supervision and Training Scale

2.5.1. Measure description

The MISTS was designed to assist in the training and supervision of therapists implementing treatments

using MI as a core element of the intervention (Madson et al., 2005). It includes two components, a behavioral count of the types of therapist responses uttered during sessions (e.g., reflection, open question) and a 16-item global rating of the quality, MI fidelity, and effectiveness of therapist interventions. The behavioral count section is completed while reviewing a session; the global ratings are completed after reviewing the session. Global ratings are assigned using a 7-point Likert-type scale on 16 aspects of MI therapy (e.g., simple and complex reflection, rolling with resistance, spirit of MI) considered central to the approach. Item-specific behavioral descriptors at Points 1, 4, and 7 were developed for each of the 16 items.

2.5.2. Psychometric properties

Madson et al. (2005) estimated the interrater reliability for the MISTS using the generalizability theory and found an overall generalizability coefficient of .79. The ICCs for the 16 individual MISTS items ranged from .41 to .81. Both convergent and discriminant validity were assessed by comparing the total score on the MISTS with the six subscales of the YACS. Support for validity was found through expected correlations between the MISTS total score and the support, goals, and 12-step as well as clinical management subscales of the YACS (Madson et al., 2005).

2.5.3. Evaluation

The MISTS shows promise for use in assessing the adherence and competence of therapists implementing MI. Attempts were made throughout the development process to focus on and retain the spirit of MI while developing the MISTS items, at the same time developing a tool that has greater utility in training and supervision. Madson et al. (2005) developed items, item-specific behavioral anchors, and a manual with the major intent to facilitate training.

Although the MISTS shows promise for use in training, supervision, and research settings, several of its limitations require revision as well as further study and preclude its use in education and training settings. Most notable is the variability in reliability estimates of individual MISTS items. Although these coefficients are considered fair to excellent according to the classification of clinical significance by Cicchetti (1994), clearly the instrument would benefit if the problematic items were stronger. For example, most of the items contained multidimensional behavioral anchors that most likely contributed to lower reliability estimates between raters. This limitation must be addressed before widespread use of the measure. The psychometrics of the MISTS has not been studied extensively. As a result, it is unclear how well the MISTS would perform outside the research environment. Therefore, further psychometric investigation of the MISTS in both research and training settings is warranted.

3. Discussion

Instruments that assess both adherence to particular forms of evidence-based treatments and competence with which therapists implement those therapies are needed for psychotherapy research and training to advance. This need is particularly important in the substance use treatment field in which there is a wide variability among the training and clinical experiences of treatment providers (Culbreth, 1999). Thus, it is encouraging to note that several efforts to establish such instruments have been undertaken with MI.

The development of these measures demonstrates a positive first step in enhancing the training, supervision, and study of MI. There are several strengths and similarities that require highlighting. Most of the measures made a concerted effort to remain true to the tenets of MI. For example, with the exception of the YACS and the MIPC, developers used the MISC as a foundation. This finding is important given the concern of several authors that researchers and practitioners may be drifting from the original intent of MI (Moyers et al., 2003; Rollnick & Miller, 1995). Another strength, specifically relating to training and supervision, is that several of these measures were developed with the intent for use in a training or supervisory environment. Supervisors have often relied on clinical expertise to provide these evaluations. Evaluation tools such as those described will eventually help trainers and supervisors facilitate skill development through evaluations that are guided by the tenets of the intervention and by empirical methods.

This review also uncovered several areas that need to be developed further to strengthen these measures before use in education and training settings. Most notable is the variability in reliability. For example, the interrater reliability of items on the two newer measures, the MITI and the MISTS, varied greatly. Given that an important aspect of these measures is an individual's ability to use them reliably to foster training, supervision, and research, these findings are concerning. However, this limitation may be explained in part by the fact that MI is a complex treatment approach in which there is continued uncertainty regarding the active ingredients in the intervention (Longabaugh, 2001; Rollnick, 2001). Nevertheless, the developers of these measures will need to address issues related to the variability in item reliability to strengthen the measures.

In addition to limitations surrounding reliability, this review identified some concerns related to how validity was estimated. Specifically, some of these measures incorporated techniques that leave questions about the validity strength of the measures. For example, Barsky and Coleman (2001) used focus groups of practitioners to develop the MIPC and found some consistency between the practitioner language and the literature language as their sole source of estimating validity. Use of a method such as this raises concern about who the professionals were in the focus group, how they were trained in MI, and how they became experts in MI.

These questions can cause speculation about what is actually being measured. Therefore, it may be helpful in further refinement and study of these measures for researchers to incorporate a variety of methods of estimating validity, such as examining the relationship of these measures to each other, to measures that are used with discriminant forms of psychotherapy, and to alternate forms of psychotherapy. Methods such as these, which are often standard in estimating validity, can significantly enhance the further refinement of the measures with the goal of strengthening their scientific base (Calsyn, 2000).

A final limitation relates to the intended use of these measures, which may complicate use in educational and training settings. For example, the YACS was developed mainly to address internal validity issues within psychotherapy clinical trials. Thus, these measures are extensive and complex. Although good measures for research use, they may have less utility in other settings. Modifications to reduce length and complexity would be needed to enhance the utility of these measures in education, training, and clinical supervision settings.

The need for evaluating clinician adherence and competence with empirically supported interventions in substance abuse treatment is growing as more evidence supporting the efficacy of interventions in facilitating behavior change develop. Evaluating clinician adherence and competence will not only help ensure that clinicians are following the tenets of an intervention appropriately but also have the potential to facilitate skill development. Although these evaluations have traditionally relied on clinical judgment, the increasing complexity of interventions combined with the call for accountability suggests that evaluators will be helped by the guidance of empirically sound evaluation tools. The measures reviewed here signify a step forward in addressing these needs. However, further refinement and study of each measure are warranted owing to limitations and unanswered questions. With scientifically sound tools, our ability to study the process of training and supervising therapists will evolve. For example, we can examine the impact of comparing therapist and supervisor ratings on skill development or the impact of therapist self-ratings using one of these measures on skill development. In addition to enhancing internal validity in treatment studies, scientifically sound measures will assist in evaluating the process of change when MI is used. Given the potential for these measures to contribute to the training (MIPC, MITI, MISTS) and study (MISC, MITI, YACS) of MI, it would behoove researchers to move forward with revising and evaluating these measures.

References

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.

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

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.

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

Campbell, T. C., Catlin, L., & Melchert, T. P. (2003). Alcohol and other drug abuse counselors' attitudes and resources for integrating research and practice. *Journal of Drug Education, 33,* 307–323.

Campbell, T. C., Daood, C., Catlin, L., & Moskovitz, A. (2005). Integration of research and practice in substance use disorder treatment: Findings from focus groups with clinicians, researchers, educators, administrators, and policy-makers. *Journal of Addictions and Offender Counseling, 26,* 4–14.

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.

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

Culbreth, J. R. (1999). Clinical supervision of substance abuse counselors: Current and preferred practices. *Journal of addictions and offender counseling, 20,* 15–25.

deJonge, J. M., Schippers, G. M., & Schaap, C. P. D. R. (2005). The Motivational Interviewing Skill Code: Reliability and a critical appraisal. *Behavioural and Cognitive Psychotherapy, 33,* 285–298.

Falender, C. A., & Shafranske, E. P. (2004). *Clinical supervision: A competency-based approach.* Washington, DC: American Psychological Association.

Institute of Medicine. (1998). *Bridging the gap between practice and research: Forging partnerships with community-based drug and alcohol treatment.* Washington, DC: National Academy Press.

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

Madson, M. B., Campbell, T. C., Barrett, D. E., Brondino, M. J., & Melchert, T. P. (2005). Development of the Motivational Interviewing Supervision and Training Scale. *Psychology of Addictive Behaviors, 19,* 303–310.

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 edition). 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: The Guilford Press.

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.

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

Orwin, R. G. (2000). Assessing program fidelity in substance abuse health services research. *Addiction*, *95*(Suppl 3), S309–S327.

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*, 235–334.

Tappin, D. M., McKay, C., McIntrye, 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.