



HHS Public Access

Author manuscript

J Subst Abuse Treat. Author manuscript; available in PMC 2017 August 02.

Published in final edited form as:

J Subst Abuse Treat. 2016 June ; 65: 36–42. doi:10.1016/j.jsat.2016.01.001.

The Motivational Interviewing Treatment Integrity Code (MITI 4): Rationale, preliminary reliability and validity

Theresa B. Moyers,

University of New Mexico, Department of Psychology, Albuquerque, New Mexico 87131,
tmoyers@unm.edu

L.N. Rowell,

University of New Mexico, Department of Psychology

Jennifer K. Manuel,

San Francisco V.A. Medical Center

Denise Ernst, and

Denise Ernst Training and Consulting

Jon M. Houck

University of New Mexico, Center on Alcoholism, Substance Abuse and Addictions

Abstract

The Motivational Interviewing Treatment Integrity Code has been revised to address new evidence-based elements of motivational interviewing (MI). This new version (MITI 4) includes new global ratings to assess clinician's attention to client language, increased rigor in assessing autonomy support and client choice, and items to evaluate the use of persuasion when giving information and advice. Method: Four undergraduate, non-professional raters were trained in the MITI and used it to review 50 audiotapes of clinicians conducting MI in actual treatments sessions. Both kappa and intraclass correlation indices were calculated for all coders, for the best rater pair and for a 20% randomly selected sample from the best rater pair. Results: Reliability across raters, with the exception of Emphasize Autonomy and % Complex Reflections, were in the good to excellent range. Reliability estimates decrease when smaller samples are used and when fewer raters contribute. Conclusion: The advantages and drawbacks of this revision are discussed including implications for research and clinical applications. The MITI 4.0 represents a reliable method for assessing the integrity of MI including both the technical and relational components of the method.

Keywords

MITI 4; Motivational Interviewing; Treatment Fidelity; Treatment Integrity; Reliability; Supervision

Introduction

Although motivational interviewing (MI) is an empirically-supported treatment with wide diffusion in a variety of clinical and healthcare settings, results of clinical trials and meta-analyses using this method yield erratic effect sizes, while findings are peppered with null results and site effects within the same trial (Burke et al., 2003; Hettema et al., 2005). One possible explanation for this lack of uniformity in outcomes concerns the integrity with which this complex clinical method is delivered. Randomized controlled trials with MI demonstrate that described interventions often contain elements that are incompatible with the method (instruction, mandatory treatment plans) or peripheral to it (decisional balance, personalized feedback) and may lack critical procedures for insuring integrity for the MI method (Madson, Loignon & Lane, 2009; Miller & Rollnick, 2014).

Several instruments have been developed to evaluate the integrity of MI in clinical trials, each with strengths and weaknesses (Madsen & Campbell, 2006). MI integrity tools initially created for research purposes are often too complex for practical supervision, such as the Motivational Interviewing Skills Code (MISC; Miller et al., 2003), the Yale Adherence and Competence Scale (YACS; Carroll et al., 2000), and the Independent Tape Rater Scale (ITRS; Ball, Martino, Covino, Morgenstern, & Carroll, 2002). Instruments created expressly for clinical supervision, including the Motivational Interviewing Supervision and Training Scale (MISTS; Madson et al., 2005) and the Motivational Interviewing Process Code (MIPC; Barsky & Coleman, 2001) are more suitable for everyday clinical practice, but typically lack counts of specific behaviors that may be especially important in conveying key aspects of MI.

With more than 260 citations in the scientific literature (Web of Science, 2015), the most commonly used tool to evaluate the fidelity of MI is the Motivational Interviewing Treatment Integrity coding system (MITI; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005). Originally created as a research tool, the MITI has proved useful in clinical settings where rigor in supervision and evaluation is needed (Manuel & Drapkin, 2015), and the manuscript describing its reliability and validity is among the top ten most frequently downloaded articles in the *Journal of Substance Abuse Treatment* (<http://www.journals.elsevier.com/journal-of-substance-abuse-treatment/most-cited-articles>). The MITI has several advantages as a fidelity measure. Chief among them is the counting of particular types of clinician behaviors (such as Questions and Reflections), which offer greater precision than simply measuring global aspects of clinician skill. Another advantage of the MITI is the rating of the clinician's expression of empathy, a core characteristic in MI and common to other psychosocial interventions. Further, the MITI has demonstrated acceptable psychometric properties across a variety of research settings (Campbell et al., 2009; Martino et al., 2008; Parsons et al., 2007; Pierson et al., 2007; Turrise et al., 2009;) and summary measures from the MITI have correlated with client outcomes in the expected direction (McCambridge, Day, Thomas & Strang, 2011; Pollack et al., 2014; Woodin, Sotskova & O'Leary, 2012).

Despite its widespread use as both an integrity measure and for clinical supervision there are notable weaknesses that have prompted a revision of the MITI. For example, client language

including language in favor of change (change talk) and in favor of not changing (sustain talk) is not represented in the MITI except obliquely in the global rating for Evocation. As greater attention has been paid to therapist's skill in evoking change talk and reducing sustain talk in MI sessions (Miller & Rose, 2009) it has become apparent that the MITI cannot inform this central element of MI practice. Similarly, the use of direct influence in the form of expert advice and information-giving is crudely measured in the MITI despite the fact that it is becoming increasingly common in MI interventions, especially in healthcare and corrections settings. More careful evaluation of these components of MI practice is warranted since they can be delivered in a manner that violates the integrity of the method.

Finally, the current MITI suffers from an inability to measure advanced or sophisticated practice in MI by an overemphasis on basic counseling skills such as asking open questions and affirming the client. While these are important foundations for MI, they are properly thought of as basic skills that might be used in a variety of humanistic interventions. More sophisticated skills that are more centrally characteristic of MI are much less carefully assessed, for example the emphasis on the client's choice about whether and how to make change. Because of this, MI is not well differentiated from other therapeutic approaches such as supportive counseling or client-centered therapy.

This article introduces the new version of the MITI that encompasses changes to address the weaknesses discussed above and provides information about its reliability. As with previous versions, the MITI 4 (Moyers, Manuel & Ernst, 2010) is an open-source instrument available for download from the CASAA website (<http://casaa.unm.edu/codinginst.html>). It yields behavior counts and global summary scores that allow the evaluation of treatment fidelity to MI in clinical trials as well as for clinical supervision purposes. With appropriate attention to its limits, the MITI 4 can be used as a supervisory or coaching tool, and some of the elements can be used for self-evaluation of MI practice.

Method

Development of the MITI 4

Major changes to the MITI coding system include: 1) developing new global ratings to assess the technical element of MI, 2) altering MI-Adherent scores to make them more specific to MI rather than general therapy skills, 3) adding a behavior count for emphasizing autonomy and choice, 4) adding behavior counts to assess the clinician's use of persuasion, collaboration and autonomy support when giving information or advice, and 5) removing the distinction between open and closed questions. The first draft of the MITI 4 was completed in May of 2012. Beta testing of this initial draft was then completed by six experienced MITI coders recruited from the Motivational Interviewing Network of Trainers (MINT) using a set of five audiotapes and transcripts from a series of DVDs (Miller, Moyers & Rollnick, 2013) designed to address the latest advancements in MI (Miller & Rollnick, 2012). The draft was revised as a result of input from this group and a new version moved forward for reliability testing with a group of five undergraduate coders in December of 2013. Results of pilot-testing with these undergraduate coders indicated that several items, though producing high reliability among expert raters, were not clear enough for non-

experts. These items included: 1) differentiation of fact-finding questions, 2) a behavior count indicating preferential responding to change talk and 3) a global rating for navigating in the direction of a desired change. These items were deleted and further changes were made to include global ratings concerning the technical elements of MI. This version, termed the MITI 4, was used for the reliability and concurrent validity analyses in this project.

As with previous versions, the MITI 4 consists of both global ratings and individual behavior counts. The MITI 3.0 had five global scores whereas the MITI 4.0 now has four (Cultivating Change talk, Softening Sustain Talk, Partnership, and Empathy), which are scored on a Likert-type scale from 1 (low) to 5 (high). In addition, the MITI 3.0 had 11 behavior counts and the MITI 4.0 now has 10: Questions, Simple Reflections, Complex Reflections, Persuade with Permission, Giving Information, Affirmations, Emphasize Autonomy, Seeking Collaboration (for example in negotiating an agenda or asking permission before giving information), Persuade and Confront (see Table 1).

Recruitment and Training of Raters

Four undergraduate raters from the University of New Mexico were recruited to work for two consecutive semesters beginning in August of 2014. These students received academic credits for their involvement in the project and worked nine hours per week for two semesters. They received 20 hours of formal training focused on the MITI 4 manual and 20 additional hours of practice with self-paced learning materials including structured coding tasks and group review of audio samples during weekly meetings. After initial training, coders worked independently, rating tapes from a library of expert-coded examples. When learner codes for these practice samples reached the “good” or “excellent” level of reliability for all items, students were approved to code actual sample tapes. Coders were masked as to when they began coding the actual study sample sessions.

Weekly group coding meetings were held for the duration of the project to prevent coder drift. Weekly meetings consisted of quizzes and learning activities, review of problematic coding and group coding of tape segments. Once work samples had been rated by a coder, those scores were considered “in the bank”, meaning they could be reviewed for mistakes and learning purposes during group meetings, but scores were never altered.

Sample of Audiotapes

The audiotaped samples were selected from the ELICIT project (Moyers et al., 2011). This parent study produced 778 audiotaped work samples of substance abuse clinicians using MI in their treatment settings, with actual clients and included baseline as well as follow-up sessions after MI training and feedback. These work samples had been previously coded with the MISC 2.5 coding system (Houck, Moyers, Miller, Glynn & Hallgren, 2010), yielding extensive information about the clinician’s proficiency with MI in each work sample. From these work samples, a group of 50 were selected with the goal of representing high, medium and low levels of MI proficiency.

Within each audiotaped work-sample a random segment of 20 minutes was selected for rating with the MITI 4. As with all previous versions of the MITI, twenty minute segments

were used to: a) allow a sample of behavior sufficient to observe relevant behaviors, without b) taxing coders to recall multiple global scales across a longer time span. All four undergraduate raters coded all 50 tapes.

Rater reliability for MITI 4

Consistent with practices in our group (Moyers et al. 2009) and with more recent recommendations (Hallgren, 2012) inter-rater reliability was assessed with 2-way mixed effects, absolute agreement, average-measures intraclass correlation (ICC) to assess the degree that coders provided consistency in their global ratings and behavior counts across tapes. This ICC method assumes fixed effects for coders and random effects for therapy sessions. This approach yields a more conservative measure of inter-rater reliability than Cronbach's alpha (Fleiss and Shrout, 1978), and is the recommended approach considering: 1) our data are ordinal, 2) more than two coders are used and 3) all coders evaluate all available samples (Hallgren, 2012). To estimate the reliability that is likely to be seen in real world coding applications, where a single coding pair is typically used, we also calculated reliability estimates using our two strongest coders. Finally, we selected a random sample of 20% of samples from the original pool of 50 ($n = 10$) and calculated the reliability estimates for the same strongest pair of coders using only those 10 tapes. This was done to investigate the approximate loss of information from samples when a reduction in the number of tapes and/or raters is practical. This situation is common in research and clinical settings where only a portion of all possible work samples is reviewed. The benchmarks for ICCs are as follows (Cicchetti & Sparrow, 1981): 0.00–0.40 = poor, 0.40–0.59 = fair, 0.60–0.74 = good, and 0.75–1.00 = excellent.

Efficiency of the MITI 4: Comparison to MISC 2.5

To evaluate the ability of the MITI 4 to capture elements of more sophisticated and labor-intensive rating systems, we used a canonical correlation. Canonical correlation is used to evaluate how much common variance is present in two different sets of variables (for example, MITI and the MISC scores).

We compared the results of our MITI coding to those obtained for the identical segments rated in a prior study (Moyers, et al., 2011) using the Motivational Interviewing Skills Code 2.5 (MISC 2.5; MISC 2.5: Houck, Moyers, Miller, Glynn, & Hallgren, 2010). The extraction of MISC data for the corresponding 20-minute segments selected for MITI 4 coding was possible because sequential coding using the CASAA Application for Coding Treatment Interactions (CACTI; Glynn, Hallgren, Houck, & Moyers, 2012) preserves both the temporal sequence of behaviors and the exact time at which behaviors occurred. MISC coding data from these segments were identified using time codes embedded in CACTI output files. Because cost-effectiveness and time efficiency are important considerations in clinical supervision and evaluation, we compared MITI 4 and MISC 2.5 behavior counts for exactly the same segments to approximate the amount of information the MITI 4 could extract from the same time segments using a comprehensive and externally valid coding system.

Prediction of Client Language with MITI 4 Global Ratings

To evaluate the ability of the new technical ratings, Cultivating Change Talk (CC) and Softening Sustain Talk (SS) to predict client language in full MI sessions, we compared the relationship between scores on these global ratings in the MITI 4 and the specific behavior count tallies of change talk and sustain talk measured in the MISC 2.5. MITI 4 raters showed high inter-rater reliability (See Table 2) and because all raters had rated all sessions, MITI global ratings for CC and SS were averaged across raters prior to comparisons with MISC ratings. Relationships between all CC and SS and MISC 2.5 counts of change talk and sustain talk were evaluated using Spearman rho (See Table 4). To assess the concurrent validity of the global ratings as “snapshots” of clinician behavior throughout the sessions, the same evaluation was performed with both the client language counts from the selected segments and from the entire session.

Results

Rater Reliability

A summary of the means, standard deviations and frequencies for MITI 4 scores can be found in Table 3. Inter-rater reliability indices for all items were in the excellent and good range using all four coders. For the strongest coder pair, scores were also in the good to excellent range with the exception of Emphasizing Autonomy and % Complex Reflections. For the 20% subsample rated by the same coder pair, ICC's were much lower, especially for items with low frequency. Intraclass correlations for summary scores for the MITI 4, typically used to evaluate clinician competence in research studies, were in the good to excellent range in all variations of tapes and coders.

Efficiency of the MITI 4.0: Comparison to MISC 2.5

Collectively, the full canonical correlation model across all functions was statistically significant using Wilks's λ , $F(100, 198.01) = 4.82$, $p < .001$. The analysis yielded five significant functions with squared canonical correlations of .936, .907, .750, .550, and .500 for each successive function. The r^2 type effect size for these variates was .468. That is, the five canonical variates for the MITI collectively explained 46.8% (about half) of the variance in the selected MISC ratings for the same segments of the sessions.

Frequency of Change and Sustain Talk

Mean frequency of CT for the full MISC coded sessions was 30.75 ($SD = 24.82$, range = 0 – 120), and for the corresponding extracted 20-minute MITI coded segments was 14.32 ($SD = 11.34$, range = 0 – 44). Mean frequency of ST for full MISC coded sessions was 9.86 ($SD = 9.49$, range = 0 – 19), and for the corresponding MITI segments it was 4.95 ($SD = 5.13$, range = 0 – 37). The distribution of these variables resulted in non-parametric comparisons for the correlations reported in Table 4.

Prediction of Client Language with MITI 4.0: Comparison of Technical Global Ratings to Client Language Counts in Full Sessions

CC showed significant correlations with total client change talk for both 20-minute segments ($\rho = .560$) and for full sessions ($\rho = .549$), indicating strong (Cohen, 1988) relationships between the global rating and the client behavior counts. This suggests that clinician scores are a relatively good index of the client's tendency to offer change talk over the entire course of the session. Scores on the SS scale were not significantly related to either change or sustain talk counts (see Table 4).

Discussion

Psychometric indices from our project indicate that the MITI 4 is a reliable measure of proficiency in MI practice as we have defined it. Specifically, it yields inter-rater reliabilities in the good to excellent range for both individual items and summary scores using non-expert undergraduate coders. Even when interrater reliability is estimated from a subset of tapes ($n = 10$) or a reduced number of coders ($n = 2$), reliabilities generally remain in the good to excellent range. This indicates that the MITI 4.0 is robust enough to withstand the typical constraints imposed by clinical and research settings, where only a portion of the available worksamples can be reviewed. With regard to the concurrent validity of the MITI, we were able to capture nearly half of the information that would be generated using more labor-intensive rating tools for the same time segments using the MISC. This provides a cost-effectiveness advantage in situations where time and resources for assessing quality of MI are limited.

One improvement in the MITI 4 is apparent in the relationship between the Cultivating Change Talk global rating and the MISC change talk ratings from both 20 minutes segments and full session recordings. This suggests that this gestalt rating of a relatively brief segment of behavior is a useful indicator of a clinician's style of interaction regarding change talk. Because of the recent emphasis on attention to change talk as a technical element of MI and as a potential causal mechanism for this method (Miller & Rose, 2009; Resnicow, Gobat & Naar, 2015), it is encouraging to find that objective ratings of a relatively brief slice of clinical behavior can provide information about the therapist's overall skill in this area over the course of the session.

The same is not true for the Softening Sustain Talk global rating, which bears little relationship to the number of sustain talk statements in the longer and more carefully coded sessions. Ideally, higher scores on the SS global ought to be associated with lower frequency of sustain talk. This is of particular concern since reductions of sustain talk appear to replace change talk as a predictor of outcome in adolescent populations or those settings where legal mandates for behavior change are common (Apodaca et al., 2014; Baer, et al., 2008; D'Amico et al., 2015; Gaume, 2015, Magill, Gaume & Apodaca, 2014).

The relatively low levels of sustain talk in these therapy samples may have obscured a relationship between the SS scale and frequency counts in the longer sessions. The absence of sustain talk is common in clinical encounters, and the reasons for low levels of sustain talk can be quite different in their clinical implications. For example, sustain talk might be

absent because the client is not experiencing or verbalizing any reasons to cling to the current problematic behavior – then the clinician’s task is to consolidate, more than evoke, reasons to change. However, a lack of sustain talk can also be observed when the clinician is skillful in using MI and diverts the potential for an unproductive focus on the status quo. This is a common snarl in prevention research - the problem of not being able to know what you have prevented. To address this we added a decision-rule for the SS scale indicating that in the absence of sustain talk the clinician should receive a high rating (either a 4 or 5). This decision rule was used during the ratings for this project, and yielded good reliability among raters. The higher score was selected as a default both to increase reliability and to avoid inadvertently penalizing clinicians with low scores when clients are not actively “pushing back” against change. Initial feedback from beta-testers and clinical supervisors has convinced us that the SS scale is a valuable addition to supervision in MI, as well as a possible indicator of outcomes in adolescent and young adult populations, so we have elected to keep it despite the logistical concerns noted above.

Advantages of the MITI 4 in both research and clinical settings are noteworthy. First, the addition of the global scales for Cultivating Change Talk and Softening Sustain Talk allow supervisors and coaches in clinical settings to increase the attention that learners pay to the skills needed for influencing client language in MI. Because the clinician’s approach to client language is numerically “graded” in these global ratings, feedback can be gleaned for guiding improvement in an incremental fashion.

Second, the problem of inflating MI proficiency by overreliance on basic counseling skills has been addressed in the MITI 4 by emphasizing behaviors that are more central to the unique aspects of this method. Chief among these are emphasizing autonomy, seeking collaboration, and affirmation. Each of these critical components of MI practice have been defined in a way that allows them to be counted as specific behaviors, and relatively conservative criteria for rating them have been developed so as to avoid including clinician behaviors that are more generally sympathetic, encouraging or kind. When practiced well, MI is different from client-centered therapy and it is different from the general warmth and encouragement that is desirable in most other clinical approaches (Miller & Rollnick, 2013). This relatively conservative evaluation of MI Consistent behaviors is intended to differentiate MI from other treatment approaches to which it is often compared. It may also have the effect of raising the bar for the clinical practice of MI, particularly in situations where the acquisition of basic counseling skills has erroneously been seen as evidence of proficiency in MI.

A third advantage of the MITI 4 is the inclusion of a category that measures the use of explicit influence. The behavior count of Persuade with Permission has been controversial, since overt persuasion has typically been seen as antithetical to MI. However, as MI has disseminated to settings in which behavior change is addressed by many types of interventionists, the use of direct influence has been much more common. Examples include healthcare settings, where information-giving and advice are necessary components of good patient care (Steinberg & Miller, 2013) and corrections settings, where direct persuasion is commonly combined with MI (Clark, Walters, Gingerich & Meltzer, 2006). Indeed, the practice of avoiding explicit influence in favor of strict and uncompromised autonomy

support may *only* be practical in conventional psychotherapeutic settings, whereas MI is most commonly used *outside* the traditional therapy hour. Recognizing that MI is often combined with overt strategies for influence on the part of the interviewer, the category of Persuade with Permission is designed to capture the skilled and careful use of persuasion in a manner that does not negate the collaboration and support of autonomy that is central to MI. Direct persuasion (without explicit links to autonomy and collaboration) is still rated as a behavior that is inconsistent with the practice of MI. We are optimistic that the addition of the Persuade with Permission behavior count, as well as the careful assessment of how information is given, will allow these features of common MI variations to be evaluated more carefully. The careful measuring of persuasion and information-giving, used both well and inappropriately, might add some light to the confusing findings of MI effectiveness across diverse settings.

A final advantage of the MITI 4 is its friendliness to raters who are not experts in MI. We intentionally designed the tool so that it would not require clinical proficiency in MI. Using expert raters may actually confer a disadvantage since non-clinicians may be less likely to deliberate about all the possible meanings and intentions of what clinicians say, which sometimes bedevils MI experts and can lower their ability to rate reliably with others. This is consistent with prior behavior coding research which has shown excellent inter-rater reliability while largely employing undergraduate students and non-clinical psychology graduate students as raters (Amrhein et al. 2003; Houck, Moyers, & Tesche 2013; Miller et al. 2004; Moyers et al. 2005; Moyers & Martin 2006, Moyers & Houck, 2010).

Several drawbacks to the MITI 4 are also apparent, and are common to behavioral rating systems more generally (Hoyt, 2000). Although our interrater reliability indices are acceptable for a therapy evaluation tool, and compare favorably with other MI fidelity measures, a substantial portion of the judgments made by raters in this instrument cannot be attributed to the scoring system itself. This is likely due to idiosyncratic tendencies of the raters and to random error. The presence of both rater bias (for example a halo effect for one rater but not another) as well as random error should lead to caution when interpreting scores from the MITI 4 (or any other rating system) even when reliability is in an acceptable range.

Rater bias can be especially problematic when the MITI 4 is used as both a supervision tool and a measure of treatment integrity within the same research study. There is evidence to indicate that supervisors and therapists are more likely than objective observers to rate sessions positively (Martino, Ball, Nich, Frankforter & Carroll, 2009) for sessions incorporating MI. In our experience, supervisors working closely with clinicians to implement MI with fidelity as a clinical trial proceeds are likely to be biased in a manner that compromises their ability to further evaluate those same clinicians objectively. For example, they may be inclined to see the therapist's skills in a more favorable light, since they are comparing them to a previous level of skill. We conclude that it is not advisable to use supervisors, who have a relationship that is valuable to the supervisory process, to also serve as the raters of clinician performance when objective evaluation of treatment integrity is needed. Specifically, we recommend that MITI 4 ratings made by supervisors during the course of a clinical trial should not be reported as measures of objective treatment fidelity.

In contrast, in non-research settings such as supervision or coaching, the MITI 4 may be a useful tool for feedback even when the rating is biased by a consulting relationship (Schoo, Lawn, Rudnik & Litt, 2015). Using the MITI 4 to evaluate practice provides a degree of structure that is likely to be helpful in shaping clinician skill even when the results are presented in a manner that is not strictly objective. Supervisors can certainly choose to present only certain results from their evaluations with the MITI 4 or to use only some of the items, with the overall goal of scaffolding the learner experience toward success and confidence. Users of the MITI may choose whether it serves as a learning tool or as an objective integrity measure, as long as the same rater is not completing both tasks.

As with many behavioral ratings systems, our data show that reductions in the number of raters used or the percentage of samples reviewed from the total universe available results in lower reliability estimates. Our 20%-sample yielded some very low agreement estimates, but this may have been due to the low frequency of some of the items (e.g. Persuade with Permission and Seek). When items occur infrequently any discrepancy between raters is magnified. Better reliability estimates are more likely when the frequency of the rated items is higher (for example, when more samples are included and therefore low frequency items increase). This highlights the tension between comprehensiveness and cost-savings that lies at the heart of quality assurance review of work samples.

Another drawback of our study is the small sample size of coded tapes and the fact that they came from only one parent trial. Our data represent only a preliminary assessment of the psychometric properties of the MITI and more research with a larger and more varied data set is a logical next step.

A final drawback to the MITI 4 concerns the issue of validity. Clearly, more research is needed to investigate whether the MITI is associated with MI outcomes. With regard to the ability to properly capture the therapist's ability to facilitate change talk, all four of the global ratings correlate with change talk, rather than simply the one scale intended to evaluate it (CC). We believe this is because the ability to encourage change talk is a latent construct and therefore composed of many different behaviors, only some of which are delivered through the explicit attention from the therapist that is captured in the CC scale.

Other concerns about the MITI 4 focus on the use of a random sample from therapy session, and how well results from the sample characterize the full session. Theoretically, randomly selected portions should be generalizable to the entire therapy session, but this will depend on the amount of variability present in the therapy hour. If different kinds of interactions are included in a single session, for example when MI is combined with assessment or skills training, the sample chosen will be less accurate as an estimate of the quality of MI. There is some reason to believe that MITI scores selected from these kinds of eclectic sessions can be misleading regarding the quality of the MI offered (Moyers & Houck, 2014). This distortion can be addressed by pre-reviewing sessions to exclude portions that focus on strategies falling outside the central core of MI, or by augmenting the MITI to account for such elements as education within a therapy session, when a more comprehensive view of integrity is desired.

An even larger question concerns the face validity of the measure - in other words, is the MITI 4 actually capturing relevant elements of MI? The expert opinion of three authors (D.E., J.K.M. and T.M.) was used to create and edit many of the items, meaning that the MITI 4 represents a particular view of the ideal practice of MI and one which is unlikely to have universal approval among MI practitioners and researchers. As with all other therapy methods, there is no errorless gold standard for evaluating motivational interviewing. Any rating instrument for a treatment, including previous and current versions of the MITI, is likely to be biased in accordance with the views of the authors and should be considered in that light. No one is the ultimate expert on motivational interviewing and it is data that should shape our practice of it. The MITI is a tool that is useful in gathering data to inform MI, not one that ultimately defines MI.

The shortcomings discussed make it clear that the MITI 4 should not be used as the only source of information when high-stakes decisions are being made. For example, when evaluating therapists in a clinical trial, when recommending remediation for front-line clinicians, or for disciplinary action of any kind, multiple sources of information are indispensable. The MITI 4 measures only a limited sample of clinician behavior in a particular context and should not be generalized more broadly to the entire body of the clinician's work. Most importantly, the MITI cannot provide information about characteristics of the client, which are likely to impact clinician skillfulness when MI is used (Imel, Baer, Martino, Ball & Carroll, 2011; Moyers, Houck, Rice, Longabaugh & Miller, in press).

In conclusion, this revision of the MITI addresses several important modifications to MI practice in the last decade. It yields reliable and valid indicators of MI practice compared to previous versions of the MITI and is friendly to both researchers and clinicians. As with all objective behavioral coding systems, revisions can be expected as data are generated and subtleties explored.

References

- Allen JP, Mattson ME, Miller WR, Tonigan JS, Connors GJ, Rychtarik RG, Townsend M. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*. 1997; 58(1):7–29. <http://dx.doi.org/10.15288/jsa.1997.58.7>. [PubMed: 8979210]
- Barsky A, Coleman H. Evaluating skill acquisition in motivational interviewing: the development of an instrument to measure practice skills. *Journal of Drug Education*. 2001; 31(1):69–82. <http://dx.doi.org/10.2190/KN79-B5BU-E55K-W4RK>. [PubMed: 11338966]
- Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology*. 2003; 71(5):843. <http://dx.doi.org/10.1037/0022-006X.71.5.843>. [PubMed: 14516234]
- Campbell MK, Carr C, DeVellis B, Switzer B, Biddle A, Amamoo MA, Sandler R. A randomized trial of tailoring and motivational interviewing to promote fruit and vegetable consumption for cancer prevention and control. *Annals of Behavioral Medicine*. 2009; 38(2):71–85. <http://dx.doi.org/10.1007/s12160-009-9140-5>. [PubMed: 20012809]
- Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and alcohol dependence*. 2000; 57(3):225–238. [PubMed: 10661673]

- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*. 1981; 86:127–137. Retrieved from <http://psycnet.apa.org/psycinfo/1982-00095-001>. [PubMed: 7315877]
- Clark MD, Walters S, Gingerich R, Meltzer M. Motivational interviewing for probation officers: Tipping the balance toward change. *Federal Probation*. 2006; 70(1):38–44.
- Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology*. 1963; 16:137–163. <http://dx.doi.org/10.1111/j.2044-8317.1963.tb00206.x>.
- Deci EL, Ryan RM. Self-determination theory in healthcare and its relations to motivational interviewing: A few comments. *International Journal of Behavioral Nutrition and Physical Activity*. 2012; 9:24. <http://dx.doi.org/10.1186/1479-5868-9-24>. [PubMed: 22385839]
- Gaume, J. Motivational Interviewing Technical and Relational Skills, Change Talk, and Alcohol Outcomes - A Moderated Mediation Analysis. Poster presented at the 38th Annual Meeting of the Research Society on Alcoholism; San Antonio, TX. 2015 Jun. Abstract retrieved from: <http://www.rsoa.org/2015meet-indexPre1.htm>
- Glynn LH, Hallgren KA, Houck JM, Moyers TB. CACTI: Free, Open-Source Software for the Sequential Coding of Behavioral Interactions. *PLoS ONE*. 2012; 7(7):e39740. <http://dx.doi.org/10.1371/journal.pone.0039740>. [PubMed: 22815713]
- Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*. 2012; 8(1):23. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/>. [PubMed: 22833776]
- Hettema J, Steele J, Miller WR. Motivational interviewing. *Annual Review of Clinical Psychology*. 2005; 1:91–111. <http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143833>.
- Holsclaw, T., Hallgren, KA., Steyvers, M., Smyth, P., Atkins, DC. Measurement Error and Outcome Distributions: Methodological Issues in Regression Analyses of Behavioral Coding Data. *Psychology of Addictive Behaviors*. 2015. <http://dx.doi.org/10.1037/adb0000091>
- Houck JM, Hunter SB, Benson JG, Cochrum LL, Rowell LN, D'Amico EJ. Temporal variation in facilitator and client behavior during group motivational interviewing sessions. *Psychology of Addictive Behaviors*. (in press).
- Houck, JM., Moyers, TB., Miller, WR., Glynn, LH., Hallgren, KA. Motivational Interviewing Skill Code (MISC) version 2.5. 2010. Available from <http://casaa.unm.edu/download/misc25.pdf>
- Houck JM, Moyers TB, Tesche CD. Through a glass darkly: Some insights on change talk via magnetoencephalography. *Psychology of Addictive Behaviors*. 2013; 27(2):489. <http://dx.doi.org/10.1037/a0029896>. [PubMed: 22946856]
- Hoyt WT. Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*. 2000; 5(1):64–86. DOI: IO.I037//I082-9S9X.5 1,64. [PubMed: 10937323]
- Imel ZE, Baer JS, Martino S, Ball SA, Carroll KM. Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence*. 2011; 115(3): 229–236. <http://dx.doi.org/10.1016/j.drugalcdep.2010.11.010>. [PubMed: 21227601]
- Lundahl B, Moleni T, Burke BL, Butters R, Tollefson D, Butler C, Rollnick S. Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. *Patient Education and Counseling*. 2013; 93(2):157–168. [PubMed: 24001658]
- Magill M, Gaume J, Apodaca TR. The technical hypothesis of motivational interviewing: A meta-analysis of MI's key causal model. *Journal of Consulting and Clinical Psychology*. 2014; 82(6): 973–983. <http://dx.doi.org/10.1037/a0036833>. [PubMed: 24841862]
- Madson MB, Campbell TC. Measures of fidelity in motivational enhancement: A systematic review. *Journal of Substance Abuse Treatment*. 2006; 31(1):67–73. <http://dx.doi.org/10.1016/j.jsat.2006.03.010>. [PubMed: 16814012]
- Madson MB, Campbell TC, Barrett DE, Brondino MJ, Melchert TP. Development of the motivational interviewing supervision and training scale. *Psychology of Addictive Behaviors*. 2005; 19(3):303. <http://dx.doi.org/10.1037/0893-164X.19.3.303>. [PubMed: 16187810]
- Madson MB, Loignon AC, Lane C. Training in motivational interviewing: a systematic review. *Journal of Substance Abuse Treatment*. 2009; 36(1):101–109. [PubMed: 18657936]

- Manuel, JK., Drapkin, ML. Department of Veteran's Affairs Motivational Interviewing and Motivational Enhancement Training Programs. Presented at the Motivational Interviewing Network of Trainers Forum; Atlanta, Georgia. October 7–8, 2014; 2014.
- Martino, S. MIA-STEP: Motivational Interviewing Assessment: Supervisory Tools for Enhancing Proficiency. Northwest Frontier Addiction Technology Transfer Center; 2006. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2692429/>
- Martino S, Ball SA, Nich C, Frankforter TL, Carroll KM. Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence*. 2008; 96(1):37–48. <http://dx.doi.org/10.1016/j.drugalcdep.2008.01.020>. [PubMed: 18328638]
- Martino S, Ball S, Nich C, Frankforter TL, Carroll K. Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors and observers. *Psychotherapy Research*. 2009; 19(2):181–193. DOI: 10.1080/10503300802688460 [PubMed: 19396649]
- Martino S, Gallon S, Ball SA, Carroll KM. A step forward in teaching addiction counselors how to supervise motivational interviewing using a clinical trials training approach. *Journal of Teaching in the Addictions*. 2008; 6(2):39–67. <http://dx.doi.org/10.1080/15332700802127946>.
- McCambridge J, Day M, Thomas BA, Strang J. Fidelity to motivational interviewing and subsequent cannabis cessation among adolescents. *Addictive Behaviors*. 2011; 36(7):749–754. <http://dx.doi.org/10.1016/j.addbeh.2011.03.002>. [PubMed: 21440994]
- Miller, WR., Moyers, TB., Rollnick, S. Motivational interviewing: Helping people change. (DVD Based on Motivational Interviewing 3rd edition). Carson City, NV: The Change Companies; 2013. www.changecompanies.net/motivational_interviewing.php
- Miller, WR., Rollnick, S. Core interviewing skills. In: Miller, WR., Rollnick, S., editors. *Motivational interviewing: Helping people change*. New York: Guilford Press; 2013. p. 62-73.
- Miller WR, Rollnick S. The effectiveness and ineffectiveness of complex behavioral interventions: impact of treatment fidelity. *Contemporary Clinical Trials*. 2014; 37(2):234–241. <http://dx.doi.org/10.1016/j.cct.2014.1.005>. [PubMed: 24469237]
- Miller WR, Rose GS. Toward a theory of motivational interviewing. *American Psychologist*. 2009; 64(6):527–537. <http://dx.doi.org/10.1037/a0016830>. [PubMed: 19739882]
- Most Cited Journal of Substance Abuse Treatment Articles Published since 2010. 2015 Oct 30. Retrieved from <http://www.journals.elsevier.com/journal-of-substance-abuse-treatment/most-cited-articles/>
- Moyers, TB. Testing Theory Based Training in Motivational Interviewing. S. Martino (Chair) Training Motivational Interviewing: New Directions in Research; Symposium presented at Association of Psychological Science; New York, NY. 2015 May.
- Moyers TB, Houck JM, Rice SL, Longabaugh R, Miller WR. Therapist empathy, Combined Behavioral Intervention and alcohol outcomes in the COMBINE Research Project. *Journal of Consulting and Clinical Psychology*. 2015 in press.
- Moyers TB, Houck JM, Glynn LH, Manuel JK. Can specialized training teach clinicians to recognize, reinforce, and elicit client language in motivational interviewing? *Alcoholism: Clinical and Experimental Research*. 2011; 35(S1):296. (Abstract).
- Moyers TB, Martin T, Manuel JK, Hendrickson SML, Miller WR. Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*. 2005; 28:19–26. <http://dx.doi.org/10.1016/j.jsat.2004.11.001>. [PubMed: 15723728]
- Moyers, TB., Manuel, JK., Ernst, DA. Motivational Interviewing Treatment Integrity Code 4.2.1. 2010. Retrieved from http://casaa.unm.edu/download/MITI4_2.pdf
- O'Halloran P, Blackstock F, Shields N, Holland A, Iles R, Kingsley M, Taylor NF. Motivational interviewing to increase physical activity in people with chronic health conditions: a systematic review and meta-analysis. *Clinical Rehabilitation*. 2014; 28:1159–1171. <http://dx.doi.org/10.1177/0269215514536210>. [PubMed: 24942478]
- Parsons JT, Golub SA, Rosof E, Holder C. Motivational interviewing and cognitive-behavioral intervention to improve HIV medication adherence among hazardous drinkers: a randomized controlled trial. *Journal of acquired immune deficiency syndromes*. 2007; 46(4):443. <http://dx.doi.org/10.1097/QAI.0b013e318158a461>. [PubMed: 18077833]

- Pollak KI, Alexander SC, Tulsy JA, Lyna P, Coffman CJ, Dolor RJ, Østbye T. Physician empathy and listening: associations with patient satisfaction and autonomy. *The Journal of the American Board of Family Medicine*. 2011; 24(6):665–672. <http://dx.doi.org/10.3122/jabfm.2011.06.110025>. [PubMed: 22086809]
- Pollak KI, Coffman CJ, Alexander SC, Østbye T, Lyna P, Tulsy JA, Bravender T. Weight's up? Predictors of weight-related communication during primary care visits with overweight adolescents. *Patient education and counseling*. 2014; 96(3):327–332. <http://dx.doi.org/10.1016/j.pec.2014.07.025>. [PubMed: 25130793]
- Resnicow K, Gobat N, Naar S. Intensifying and igniting change talk in motivational interviewing: A theoretical and practical framework. *The European Health Psychologist*. 2015; 17(3):102–110.
- Resnicow K, McMaster F. Motivational interviewing: Moving from why to how with autonomy support. *International Journal of Behavioral Nutrition and Physical Activity*. 2012; 9(19) <http://dx.doi.org/10.1186/1479-5868-9-19>.
- Schoo AM, Lawn S, Rudnik E, Litt JC. Teaching health science students foundation motivational interviewing skills: Use of motivational interviewing treatment integrity and self-reflection to approach transformative learning. *BMC Medical Education*. 2015; 15:228.doi: 10.1186/s12909-015-0512-1 [PubMed: 26689193]
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86(2):420. <http://dx.doi.org/10.1037/0033-2909.86.2.420>. [PubMed: 18839484]
- Steinberg, MP., Miller, WR. Offering information and advice. In: Steinberg, M., Miller, WR., editors. *Motivational Interviewing in Diabetes Care*. New York: Guilford Press; 2015. p. 48-58.
- Tabachnick, LS., Fidell, BG. *Using Multivariate Statistics*. Boston, Mass: Allyn & Bacon; 2006.
- Turrisi R, Larimer ME, Mallett KA, Kilmer JR, Ray AE, Mastroleo NR, Montoya H. A randomized clinical trial evaluating a combined alcohol intervention for high-risk college students. *Journal of Studies on Alcohol and Drugs*. 2009; 70(4):555. <http://dx.doi.org/10.15288/jsad.2009.70.555>. [PubMed: 19515296]
- Vader AM, Walters ST, Prabhu GC, Houck JM, Field CA. The language of motivational interviewing and feedback: counselor language, client language, and client drinking outcomes. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*. 2010; 24(2): 190–197. <http://doi.org/10.1037/a0018749>. [PubMed: 20565145]
- Web of Science. [Retrieved July 28, 2015] ISI Web of Knowledge Platform. Thomson Reuters. 2015. from <http://apps.webofknowledge.com/>
- Woodin EM, Sotskova A, O'Leary KD. Do motivational interviewing behaviors predict reductions in partner aggression for men and women? *Behaviour Research and Therapy*. 2012; 50:79–84. <http://dx.doi.org/10.1016/j.brat.2011.11.001>. [PubMed: 22119133]

Table 1

Description of MITI codes.

MITI Code	Brief Description
Globals	
Cultivating Change Talk (CC)	Encourages the client's own language in favor of the change goal and confidence for making that change.
Softening Sustain Talk (SS)	Avoids a focus on the reasons against changing or on maintaining the status quo.
Partnership (P)	Conveys an understanding that expertise and wisdom about change reside mostly within the client.
Empathy (E)	Understands or makes an effort to grasp the client's perspective and experience.
Behavior Counts	
Giving Information (GI)	Gives information, educates, provides feedback, or expresses a professional opinion without persuading, advising, or warning.
Questions (Q)	Questions (open or closed).
Simple Reflection (SR)	Reflects a client's statement with little or no added meaning or emphasis.
Complex Reflection (CR)	Reflects a client's statement with added meaning or emphasis.
Affirm (AF)	States something positive about the client's strengths, efforts, intentions, or worth.
Emphasize Autonomy (EA)	Highlights a client's sense of control, freedom of choice, personal autonomy, ability, and obligation about change.
Confront (C)	Directly and unambiguously disagreeing, arguing, correcting, shaming, blaming, criticizing, labeling, warning, moralizing, ridiculing, or questioning a client's honesty.
Seek Collaboration (Seek)	Attempts to share power or acknowledge the expertise of a client.
Persuade with Permission (PwP)	Emphasis on collaboration or autonomy support while using direct influence.
Persuasion (Per)	Overt attempts to change a client's opinions, attitudes, or behaviors using tools such as logic, compelling arguments, self-disclosure, facts, biased information, advice, suggestions, tips, opinions, or solutions to problems.
Summary Measures	
Total MI Non Adherent (MINA)	MI-Non-adherent Behaviors= (Total Per) + (Total CO)
Total MI Adherent (MIA)	MI-Adherent Behaviors= (Total EA) + (Total Seek) + (Total AF)
Reflection to Question Ratio (R:Q)	Reflections to Questions Ratio= (Total Reflections)/(Total Questions)
Relational	Relational= [(Partnership) + (Empathy)]/2
Technical	Relational= [(Cultivating) + (Softening)]/2
%CR	Percent Complex Reflections= CR/(SR + CR)

Table 2

ICC for MITI Globals and Behavior Counts.

MITI Code	ICC-4 coders (N=50)	ICC-2 best coders (N=50)	ICC-2 best coders (N=10)
Globals			
Cultivating Change Talk	.910 ^{***}	.862 ^{***}	.953 ^{***}
Softening Sustain Talk	.883 ^{***}	.774 ^{***}	.911 ^{***}
Partnership	.868 ^{***}	.786 ^{***}	.643 ^{**}
Empathy	.886 ^{***}	.799 ^{***}	.787 ^{***}
Behavior Counts			
Giving Information	.901 ^{***}	.715 ^{**}	.729 ^{**}
Questions	.970 ^{***}	.876 ^{***}	.895 ^{***}
Simple Reflection	.930 ^{***}	.848 ^{***}	.587 [*]
Complex Reflection	.908 ^{***}	.714 ^{**}	.562 [*]
Affirm	.885 ^{***}	.758 ^{***}	.716 ^{**}
Emphasize Autonomy	.868 ^{***}	.577 [*]	.062
Confront	.852 ^{***}	.665 ^{**}	.815 ^{***}
Seek Collaboration	.839 ^{***}	.828 ^{***}	.640 ^{**}
Persuade with Permission	.816 ^{***}	.643 ^{**}	.216
Persuade	.908 ^{***}	.786 ^{***}	.725 ^{**}
Summary Measures			
Total MI Non-Adherent	.885 ^{***}	.741 ^{**}	.916 ^{***}
Total MI Adherent	.852 ^{***}	.778 ^{***}	.609 ^{**}
Reflection:Question	.930 ^{***}	.919 ^{***}	.590 [*]
Technical	.655 ^{**}	.844 ^{***}	.764 ^{***}
Relational	.660 ^{**}	.835 ^{***}	.511 [*]
%Complex Reflection	.732 ^{**}	.534 [*]	.580 [*]

Note.

* = Fair reliability;

** = Good reliability;

*** = Excellent (Cicchetti, 1994)

Table 3

Frequencies, Means and Standard Deviations for MITI Globals and Behavior Counts.

MITI Code	4 Coders N=50		2 Coders N=10	
	Mean	SD	Mean	SD
Globals				
CC	3.0	1.2	2.7	1.3
SS	2.8	1.1	2.7	1.3
Partnership	3.1	1.2	2.5	1.1
Empathy	3.2	1.1	2.7	1.2
Behavior Counts				
GI	2.5	2.5	4.3	5.3
Question	21.5	11.4	24.4	14.7
Simple Reflection	10.7	6.1	11.1	7.8
Complex reflection	3.8	3.2	4.2	4.5
Affirm	0.9	1.1	1.1	1.2
EA	0.3	0.6	0.5	0.6
CO	1.0	1.6	2.0	2.6
Seek	0.4	0.8	0.2	0.4
PwP	0.3	0.5	0.3	0.4
Per	2.4	3.0	5.5	4.9
Summary Measures				
Total MINA	3.5	4.7	6.9	6.2
Total MIA	1.6	2.0	1.9	1.8
Ratio of Reflections to Questions	0.9	0.7	1.0	1.0
Technical	4.4	2.3	3.9	2.4
Relational	4.7	2.2	2.5	1.2

Table 4

Spearman rho correlations of MITI global ratings with change talk (CT) and sustain talk (ST) counts

Change language in MISC	CC	SS	Partnership	Empathy
Session CT	.549 ^{***}	.522 ^{***}	.518 ^{***}	.510 ^{***}
Session ST	.093	.092	.073	.090
Session %CT	.374 [*]	.341 [*]	.292	.278
Segment CT	.560 ^{***}	.520 ^{***}	.527 ^{***}	.510 ^{***}
Segment ST	.079	.087	.110	.101
Segment %CT	.384 [*]	.330 [*]	.307 [*]	.285

Note. N=46.

* = p < .05;

** = p < .01;

*** = p < .001.

%CT = Percent CT (i.e., CT/(CT+ST))

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript