

ASSESSING THE INTEGRITY OF MOTIVATIONAL INTERVIEWING INTERVENTIONS: RELIABILITY OF THE MOTIVATIONAL INTERVIEWING SKILLS CODE

Theresa Moyers and Tim Martin

University of New Mexico, Albuquerque, USA

Delwyn Catley, Kari Jo Harris, Jasjit S. Ahluwalia

University of Kansas Medical Center, USA

Abstract. The motivational interviewing skills code (MISC) was used to review 86 audiotaped interactions between clinicians and patients participating in a smoking cessation intervention. Intraclass correlations (ICCs) were completed for two of the MISC elements: global evaluations and behavior counts. Results indicate 75% of the global ratings yielded ICCs in the good to excellent range, while only 44% of the behavior counts yielded this level of accuracy. Adherence scores were created to form overall ratings of clinician adherence to using motivational interviewing and 80% of these competence measures yielded ICCs in the good to excellent range. Specific recommendations regarding modifications for the MISC are suggested based on the data presented.

Keywords: Motivational interviewing, reliability, treatment integrity.

Introduction

Motivational interviewing (MI) is a therapeutic method designed to increase readiness to change in situations where ambivalence stymies adaptive life choices. Such ambivalence commonly occurs with problems such as addiction, weight change and response to chronic health conditions. Motivational interviewing incorporates client-centered, cognitive-behavioural and humanistic traditions, but it is unique in focusing in helping clients to resolve mixed feelings about change, with the assumption that change itself can be relatively straightforward once such ambivalence is reduced. Reviews of randomized, controlled trials (Burke, Arkowitz, & Dunn, 2002; Noonan & Moyers, 1997) indicate that MI (combined with objective feedback) is helpful both as a stand-alone treatment and as a prelude to other interventions. Despite such promising evidence of usefulness, questions remain about the

Reprint requests to Theresa B. Moyers, Center on Alcoholism Substance Abuse and Addictions, University of New Mexico, 2650 Yale SE, Albuquerque, NM 87106, USA. E-mail: tmoyers@unm.edu

© 2003 British Association for Behavioural and Cognitive Psychotherapies

accurate quantification of this method. Studies purporting to use MI sometimes contain descriptions of interactions with patients that blatantly violate the spirit of this method (Noonan & Moyers, 1997; Rollnick & Miller, 1995). Even when treatment descriptions are consistent with this method, very few randomized, controlled studies of MI include any integrity measures to ensure that they are delivered as described (Burke et al., 2002).

In response to these needs, the Motivational Interviewing Skills Code (MISC) was developed to quantify motivational interviewing during clinical interactions in which a specific behavior change is desirable (Miller & Mount, 2001). The MISC uses three methods for quantifying therapist-client interactions, which are assessed during independent review (or “passes”) of a segment of audio or videotape.

Description of the MISC

The first pass yields global scores on qualitative dimensions of theoretical importance to motivational interviewing, similar to those developed by Truax and Carkhuff (1967) to measure the characteristics of client-centered therapists. In the MISC, clinicians are rated on the dimensions of acceptance, egalitarianism, warmth, genuineness, empathy and overall adherence to the spirit of motivational interviewing. Clients are rated on the dimensions of affect, cooperation, disclosure and engagement. The interaction between the client and therapist is evaluated on the two dimensions of collaboration and benefit.

In contrast to the qualitative or “gestalt” measures in the first pass, the second pass consists of frequency counts of specific behaviors for both clinicians and clients, comparable to those used by Chamberlain et al. (2001). These behaviors constitute an exhaustive and mutually exclusive compendium. Twenty-seven therapist behaviors and four client behaviors are counted. In addition, summary scores of therapist adherence to the MI method can also be formed by combining all therapist behaviors consistent with motivational interviewing made during the designated time period (MICO) as well as those inconsistent with motivational interviewing (MIIN).

The third pass of the MISC is a measure of the amount of time spent speaking in the session by both the clinician and client. This “talk time” is expressed as a ratio of clinician to client use of the available time for speech in the session. Time spent in silence is not measured in this ratio.

Studies using the MISC

Psychometric support for the MISC is sparse, yet a handful of studies have been reported in the literature indicating the potential usefulness of this tool. The *construct validity* of the MISC was investigated by Miller and Mount (2001) who rated 4 experts using motivational interviewing with role-played clients. Reasoning that the experts should produce scores consistent with proper use of the method, the taped segments were independently reviewed by 5 raters, using the MISC. Results indicated that experts typically exceeded a 5 (on a 7-point Likert scale) rating level for global dimensions such as acceptance and egalitarianism, achieved a Reflection to Question ratio greater than 2, and Percent Complex Reflections greater than 50%. The average MI Consistent (MICO) percentage for these experts was over 90% and Therapist Talk Time was less than 50%. The *reliability* of the MISC was evaluated in a study by Tappin et al. (2000) that evaluated the use of specially trained midwives to

help pregnant smokers reduce their smoking during home visits. Tappin et al. collapsed across all six therapist ratings in the first pass to form one measure of therapist functioning. Similarly, the four client scales were collapsed to form a single client variable and the two interaction scales were combined to form one measure of the interaction between the client and therapist. The intraclass correlation of the collapsed measures was .39, .53 and .51. Correlations for the individual behavior counts (second pass) were not reported but the adherence scores were computed and yielded ICCs of .45 for MICO behaviors and .67 for MIIN behaviors. Client behaviors were rated much more reliably, with ICCs of .77 for speech indicating change and .76 for client speech indicating resistance. The third pass was not reported.

These initial studies indicate that the MISC shows some promise for capturing important elements of both therapist and client functioning within the context of a motivational interviewing session. Important questions remain concerning the psychometric properties of this instrument, particularly in regard to the adherence scores intended to document overall therapist functioning. The next step in investigating this promising research tool would involve investigating the reliability of the MISC using a larger clinical sample of client-therapist interactions, from which reliability estimates could be derived from all individual items in the MISC subscales, as well as the adherence measures. The present study investigates the reliability of the MISC by using data from a large clinical trial implementing MI with smokers. It has the advantage of a larger number of therapists, as well as a greater total number of coded samples, than those used in previous psychometric investigations of the MISC. It provides a venue for investigating both the inter-rater reliability of all items on the Motivational Interviewing Skills Code, as well as determining whether this instrument yields any reliable overall measures of therapist competency in the use of motivational interviewing that might be suitable for use as an appraisal of treatment integrity.

Methods

Source of tapes

This study examined audiotapes from the “Kick-it at Swope” study at the University of Kansas, investigating the effectiveness of bupropion as an aid to smoking cessation among African Americans. MI was used as an adjunct to treatment for participants in both the treatment and placebo groups. At the randomization visit participants were provided with medication and received approximately 20 minutes of counseling. Eighty-nine consecutive baseline sessions were tape-recorded but due to recording problems only 86 were usable. All 86 were coded using the MISC coding system.

Counselors

Sessions were conducted by one of four African American Masters level counselors who had received training in MI. Counselor’s training included attendance at an introductory workshop on MI as well as project-specific training sessions conducted by a study consultant with expertise in using MI in research interventions. Counselors were trained to follow a session outline developed for the study, which included exploring ambivalence about quitting, preparing to quit, problem solving difficult situations, and problem solving on adher-

ence with the medication. However, because MI was not a focus of the original study, counselors were not “certified” as skilled in MI and did not receive ongoing supervision or verification of their fidelity to MI principles.

Selection and training of coders

Coders were two upper-level experimental psychology graduate students trained by the first author. Total training time involved 40 hours over 4 weeks and utilized a graduated approach whereby coders 1) initially coded tapes simultaneously with an expert, 2) then coded tapes independently (with results double-coded by the expert and differences resolved in a weekly meeting) and 3) coded independently with a random sample double-coded by the expert and differences resolved in a weekly staff meeting. None of the tapes used to provide data for this study were used as training tapes for the coders.

Process of coding

Following procedures established in the MISC coding manual (available for download at no charge at casaa.unm.edu), each coder listened to the designated 20-minute tape sample twice. The first time, the coder listened to the segment without stopping, then indicated the appropriate global ratings. The second time, the coder was permitted to turn the recorder on and off as often as needed to assign codes for each therapist and client utterance. The third (timing) pass was completed by undergraduate students. Detailed procedures for parsing utterances, determining behavior codes and completing timing calculations are found in the MISC coding manual (available free of charge at www.casaa.unm.edu).

Table 1. Global measures of therapist, client and interaction: Pass one

Measure	ICC	Category	α	r
Therapist	.3895	Poor	.5606	.421**
Acceptance				
Egalitarianism	.6758	Good	.8066	.690**
Empathy	.7652	Excellent	.8670	.769**
Genuineness	.6888	Good	.8157	.689**
Warmth	.4769	Fair	.6458	.548**
Spirit	.7912	Excellent	.8835	.791**
Client				
Affect	.3986	Poor	.5700	.399**
Cooperation	.2524	Poor	.4031	.255
Disclosure	.7326	Good	.8457	.739**
Engagement	.7583	Excellent	.8625	.764**
Interaction				
Collaboration	.5829	Fair	.7365	.626**
Benefit	.6315	Good	.7741	.675**

** $p < .01$

Reliability subsample

After all 86 tapes had been coded, a convenience sample of 50 randomly selected tapes were chosen and re-coded by the alternative coder in the pair. Coders were blind to previous ratings. Previous findings in our laboratory indicate that the third pass (Therapist/Client Talk Time) has extremely high reliability, even when using inexperienced undergraduate students. Therefore, pass three was not included in the double-coded sample.

Table 2. Behavior counts: Therapist and client

Measure	ICC	Category	α	r
Therapist	-.0441	Poor	-.0922	-.045
Advice with permission				
Advice without permission	.0303	Poor	-.0625	-.031
Affirm	.3768	Poor	.5473	.391**
Confront	.0000	Poor	.0000	0
Direct	.5680	Fair	.7245	.604**
Emphasize control	1.000	Excellent	1.000	1.000**
Facilitate	.6483	Good	.7866	.669**
Filler	.5580	Fair	.7163	.561**
Personal feedback	.1716	Poor	.2929	.193
Self-disclosure	.3153	Poor	.4794	.352*
General information	.7592	Excellent	.8631	.761**
Closed question	.9036	Excellent	.9493	.914**
Open question	.5459	Fair	.7063	.611**
Raise concern with permission	—	—	—	—
Raise concern without permission	.2070	Poor	.3430	.231
Repeat	.7641	Excellent	.8663	.774**
Repeat (affect)	.2020	Poor	.3361	.232
Rephrase	.7702	Excellent	.8702	.779**
Rephrase (affect)	.5930	Fair	.7445	.600**
Paraphrase	.6945	Good	.8197	.762**
Paraphrase (affect)	.3146	Poor	.4786	.327**
Summarize	.1509	Poor	.2622	.192
Summarize (affect)	.3786	Poor	.5493	.390**
Reframe	.7911	Excellent	.8834	.858**
Structure	.2146	Poor	.3534	.219
Support	.6473	Good	.7859	.661**
Warn	.3684	Poor	.5384	.374**
Client				
Ask	.8653	Excellent	.9278	.867**
Follow/neutral	.5761	Fair	.7310	.586**
Change	.7815	Excellent	.8774	.786**
Resist change	.5314	Fair	.6940	.540**

* $p < .05$ ** $p < .01$

Table 3. Intraclass correlations for reflections collapsed across affect

Measure	ICC	Category	α	r
Repeat	.7154	Good	.8279	.7278**
Rephrase	.8160	Excellent	.8987	.8216**
Paraphrase	.7795	Excellent	.8761	.8183**
Summarize	.2442	Poor	.3925	.2942*

* $p < .05$ ** $p < .01$

Results

Was acceptable interrater reliability achieved using the MISC?

Several measures were calculated to assess inter-rater reliability. Because of their familiarity and ease of interpretation, Pearson product moments and Cronbach's alpha were calculated for each measure. Intra-class correlation coefficients were also calculated. This more conservative statistic adjusts for change agreement between raters, as well as systematic differences between raters, and therefore is probably preferable in assessing inter-rater reliability. Cichetti (1994) has proposed a categorization system for evaluating the usefulness of ICCs in clinical instruments: below .40 = poor, .40 to .59 = fair, .69 to .74 = good, and .75 to 1.00 = excellent. ICC values for the global measures can be found in Table 1, while those for the behavior counts are found in Table 2.

We were concerned that several behaviors inconsistent with MI and therefore of clinical theoretic importance (such as Confront and Warn) failed to yield acceptable reliability scores. The correlation between behavior frequency and reliability ($r = .390$, $p = .033$) indicates that one hypothesis for low reliability for these items is that they are encountered too infrequently for coders to learn to categorize them consistently.

Another obvious problem concerned therapist reflections of client affect, which as a class proved unreliable. We therefore collapsed reflection counts across affect and recalculated the ICCs, which yielded improved reliability (Table 3).

Does this MISC yield reliable adherence scores useful for indicating therapist adherence in using motivational interviewing?

As noted previously, Miller and Mount (2001) used the MICO and MIIN scales to estimate treatment adherence and Tappin et al. (2000) used slightly different items to comprise their MICO adherence score. Using Tappin et al. (2000) formula, we found that even when ICCs for a few individual items included in the adherence measures were poor, the overall measures yielded good to excellent interrater reliability, with the exception of MI inconsistent behaviors (see Table 4).

Discussion

These results indicate that the MISC generally offers good interrater reliability in capturing the global dimensions of MI sessions. In addition, the adherence measures capturing the extent to which the therapist uses behaviors consistent with motivational interviewing can

Table 4. Adherence measures of therapist-client functioning from the MISC

Measure	ICC	Category	α	r
Therapist	.7738	Excellent	.8725	.774**
Client	.7481	Good	.8559	.748**
Interaction	.6333	Good	.7755	.683**
MI consistent	.8092	Excellent	.8945	.832**
MI inconsistent	.5107	Fair	.6761	.514**

* $p < .05$ ** $p < .01$

also be reliably measured, offering a useful integrity check for compliance with this clinical method.

Although we were able to code 75% of the individual items in the first (global) pass with fair, good or excellent reliability, this fell to 44% in the second (behavior counts) pass. The number of individual behavior counts on which the coders failed to achieve acceptable reliability indicates the possibility that revision of the MISC is needed to achieve a simpler set of behavior codes. An example would be the coding of affect, which posed a uniformly difficult task for our coders whenever they attempted to observe it within the MISC system.

Of more concern was our inability to attain acceptable inter-rater reliability on several items of specific theoretical importance to the use of motivational interviewing. Confrontation, which is an example of behavior inconsistent with motivational interviewing, yielded an ICC of 0. This means that every time one of the pair of independent coders noted a confront response in any taped sample, the other coder did not code it similarly. This is likely to have been influenced by the extremely low frequency of MI inconsistent behaviors such as confrontation, warning and giving advice within the highly trained therapist universe observed in this particular sample. For projects with homogenous therapists, particularly those also using standard protocols, specific training materials designed to illustrate uncommon but important therapist behaviors may be needed to train coders to acceptable inter-rater reliability.

A final drawback of the MISC in capturing important dimensions of motivational interviewing is the lack of information about sequential behaviors between the therapist and client. The current system of behavioral tallies in the MISC does not allow us to investigate what happens at the moment a therapist encounters specific client behaviors, such as resistance, nor how they respond to instances of client speech favoring change. Given the evidence indicating the importance of these elements in the effectiveness of motivational interviewing, the MISC clearly provides only a partial appraisal of this therapeutic method (Amrhein et al., 2000; Miller & Rollnick, 2002).

Despite the drawbacks of this study, these data encourage further use of the MISC to investigate both the integrity and process of motivational interviewing. In particular, the MISC offers promise as a treatment integrity check for clinical trials, as a tool for investigating some of the theoretical mechanisms of action for MI, and as a tool for providing immediate feedback for student therapists. Future revisions of the MISC must address the need to measure sequential information in therapist and client interactions.

References

- AMRHEIN, P. C., MILLER, W. R., YAHNE, C., PALMER, M., & FULCHER, L. (2000). *Committing language emergent from a motivational interview predicts behavioral change in drug-addicted clients*. Presented at the Ninth International Conference on Treatment of the Addictions (Ictab). September. Cape Town, South Africa.
- BURKE, B. L., ARKOWITZ, H., & DUNN, C. (2002). The efficacy of motivational interviewing and its adaptations: What we know so far. In W. R. Miller & S. Rollnick (Eds.), *Motivational interviewing: Preparing people for change* (2nd Ed.). New York: Guilford Press.
- CHAMBERLAIN, P., DAVIS, J. P., FORGATCH, M., FREY, J., PATTERSON, G. R., RAY, J., ROTHSCHILD, A., & TREMBLEY, J. (2001). *The therapy process code: A multidimensional system for observing therapist and client interactions*. Available at: www.ori.org.
- CICCHETTI, D. V. (1994). Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- MILLER, W. R., & MOUNT, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*, 29, 457–471.
- MILLER, W. R., & ROLLNICK, S. (2002). *Motivational interviewing: Preparing people to change* (2nd Ed.). New York: Guilford Press.
- NOONAN, W. C., & MOYERS, T. B. (1997). Motivational interviewing: A review. *Journal of Substance Misuse*, 2, 8–16.
- ROLLNICK, S., & MILLER, W. R. (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, 23, 325–334.
- TAPPIN, D. M., MCKAY, C., MCINTYRE, D., GILMOUR, W. H., COWAN, S., CRAWFORD, F., CURRIE, F., & LUMSDEN, M. A. (2000). A practical instrument to document the process of motivational interviewing. *Behavioural and Cognitive Psychotherapy*, 28, 17–32.
- TRUAX, C. B., & CARKHUFF, R. R. (1967). *Toward effective counseling and psychotherapy*. Chicago: Aldine.