Behavioural and Cognitive Psychotherapy, 2005, 33, 285–298
Printed in the United Kingdom doi:10.1017/S1352465804001948

The Motivational Interviewing Skill Code: Reliability and a Critical Appraisal

Jannet M. de Jonge

University of Groningen, The Netherlands

Gerard M. Schippers

Amsterdam Institute for Addiction Research, The Netherlands

Cas P. D. R. Schaap

University of Groningen, The Netherlands

Abstract. The Motivational Interviewing Skill Code (MISC) is a coding system developed to measure adherence to motivational interviewing (MI). MI is an effective clinical style used in different treatment situations. Counsellors practising MI have to follow general principles and avoid certain traps. In the present study, the content of the MISC is compared with the general principles of MI and the traps to avoid in MI. Investigation of the content validity raises some questions. All general principles are represented but the traps to avoid in MI are not fully covered. The consequences of this under-representation are shown in transcripts of a selection of well-conducted MI training sessions. The reliability of the MISC was investigated by having five independent coders code 39 MI training sessions of different counsellors. The reliability of the MISC is reasonable. The five coders agreed to a large extent on the absolute ratings but the intraclass correlations were low. Although the MISC can be a useful research tool for process research of MI, it remains a labour-intensive instrument and for teaching and practice audit development of a more simple coding system is recommended.

Keywords: Motivational interviewing, coding system, reliability, content validity.

Introduction

Motivational interviewing

Ever since Miller (Miller, 1983) first described motivational interviewing (MI) as an alternative style for working with problem drinkers, its practice has increased in treating substance abuse and in a wide variety of other mental and general health treatment situations (Miller and Rollnick, 1991). Both Dunn, DeRoo and Rivera (2001) and Noonan and Moyers (1997)

Reprint requests to Jannet M de Jonge, Department of Clinical and Developmental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. E-mail: j.m.de.jonge@ppsw.rug.nl.

systematically reviewed the effectiveness of MI in substance abuse and other behaviour domains and found it effective compared to no-treatment control groups and treatment control groups in most areas. The popularity of MI makes it worthwhile to examine in detail the way it is practised. Furthermore, studies of treatments like MI should conduct an independent check of the manipulated variable. Both the content of a treatment as well as the quality of a treatment should be examined (Kendall, Holmbeck and Verduin, 2004).

MI helps people to overcome the ambivalence that keeps them from making desired changes in their lives. Miller and Rollnick (1991, 2002) described five general principles underlying the MI style. The first principle is to express empathy with the underlying attitude of unconditional regard and acceptance. The second principle is to develop discrepancy and increase ambivalence towards change. In accordance with the third principle, the counsellor cooperates in seeking a solution for the client's ambivalence while avoiding a discussion with the client. According to the fourth principle, the counsellor must roll with the client's resistance during a session. If the resistance is influenced by the client's perception of incompetence, the counsellor has to support the client's self-efficacy. This is the fifth principle. In 2002, Miller and Rollnick presented four principles only, leaving out the third one "avoiding discussion", but without giving any further clarification as to why this was done.

By working according to these four or five general principles, a counsellor is in sympathy with the client's stage of readiness to change (DiClemente, 1991). Miller and Rollnick (1991, 2002) formulated traps to avoid in MI. The first is the *question-answer* trap in which the counsellor successively asks questions, resulting in a client giving short answers. When a counsellor *confronts* the client persuading him to change his behaviour, he evokes denial. *Labelling* the client as an alcoholic or *blaming* him for his problems also causes resistance. When a counsellor "fixes" the client's situation, prescribing solutions and pretending to be the *expert* in the client's life, the client's knowledge and ability to choose will be undervalued. A counsellor rushing into the client's alcohol abuse or other problems while the client wishes to discuss a broader range of concerns, has fallen into the *premature-focus* trap. In their second edition, Miller and Rollnick (2002) replaced the third general principle, avoiding discussion, and reformulated it as the counsellor trap of *taking sides*.

In order to examine whether MI is properly practised clinically, it is important to apply a standardized coding system for monitoring MI style and technique (Dunn et al., 2001). In general, process research needs research tools to provide information for counsellors about how to intervene with clients at different points during treatment and to develop training programs for future counsellors which include therapists, practitioners, and, in particular, non-expert helpers (Hill and Corbett, 1993; Hill and Lambert, 2004). The Motivational Interviewing Skill Code (MISC) was developed to help identify the active agents in MI. The MISC can be expected to be of help in training for MI and in testing the integrity of treatment in MI research. Data on the MISC have been reported by Miller and Mount (2001) and by Tappin et al. (2000). The latter studied the reliability of the MISC in monitoring one midwife practising MI while motivating pregnant women to stop smoking. Three coders listened to the MI sessions. Their interrater reliability on the MISC codes was low to moderate. It was suggested that more therapists and less behaviour categories may improve the interrater reliability (Tappin et al., 2000). Miller and Mount (2001) showed the discriminant validity of the MISC, using it to document change in counsellor behaviour after training in MI.

In the present study, the MISC is evaluated by a critical appraisal of its content validity and an assessment of its reliability. Five independent coders coded 39 MI role-play training

sessions, all carried out by different counsellors. The pitfalls of using the MISC as an adherence instrument for MI will be illustrated.

Motivational Interviewing skill code

The MISC (Miller, 2000) was developed by William R. Miller and colleagues at the Center for Health Research in Portland. The MISC includes three coding passes to encode audiotaped or videotaped MI sessions. The first pass consists of global judgments of the therapist, the client and the interaction between the therapist and a client. The six therapist rating scales evaluate "acceptance", "egalitarianism", "empathy", "genuineness", "warmth", and the "spirit of MI". A high score means that the counsellor shows a client-centred approach, while a low score means that the counsellor shows disrespectful, authoritarian and unfriendly behaviour. The four global client rating scales evaluate "affect", "cooperation", "disclosure", and "engagement". High scores mean that the client is emotional, open, responsive and involved, while low scores mean the opposite. The two global interaction ratings evaluate "collaboration" and "benefit". A high score on "collaboration" means that the counsellor and the client are moving together towards a motivated decision, while a low score means a competitive atmosphere during the session. A high score on "benefit" means that a client shows movement toward beneficial change, a midpoint score means a client leaves the session no different than at the start of an MI session, while a low score indicates that a client is moving away from change.

During the second pass, each counsellor and client utterance is classified into one of a set of mutually exclusive categories refined from a coding system originally developed by Chamberlain, Patterson, Reid, Kavanagh and Forgatch (1984). As Table 1 shows, the main therapist behaviour counts are advice, affirm, confront, direct, emphasize control, facilitate, filler, inform, question, raise concern, reflect, reframe, support, structure and warn. Advice can be given with and without prior permission. Inform can be subdivided into general information, personal feedback and self-disclosure by the counsellor. The counsellor can ask open and closed questions. Four types of reflections can be made: repeat, rephrase, paraphrase what the client has said, and summarize two or more client statements. All reflections can be with or without reflecting effect. The client utterances can be classified into ask, follow, resist change, and change talk.

During the third coding pass, the coder records the length of time that the counsellor and client spoke individually and the total length of time spent talking in decimal minutes.

The extent to which the MISC covers the MI principles and traps

According to the author, the MISC was developed in the first place to study in depth the interaction between a counsellor and a client, and the role that the MI style and techniques play in the client's process of change. Therefore, the feasibility of the MISC to assess the adherence of a conversation to the clinical style of MI and the content validity of the MISC concerning the do's and don'ts of MI, should be appraised separately. A key question is whether all MI principles (Miller and Rollnick, 1991), as well as the avoidance of traps in MI, are covered sufficiently in the MISC. The first thing to be noticed is that four of the five global first pass therapist ratings ("acceptance", "empathy", "genuineness", and "warmth") refer to only one principle of MI: expressing empathy. The first pass rating, *spirit of MI*, although referring to all principles, is general in itself, leaving the other principles not well represented in these

Table 1. The Motivational Interviewing Skill Code: second pass behaviour counts

Therapist behaviour counts	Subclassification	Abbreviation
Advice	with permission	AD+
Advice	without permission	AD−
Affirm		AF
Confront		CO
Direct		DI
Emphasize control		EC
Facilitate		FA
Filler		FI
Inform	general information	GI
Inform	personal feedback	PF
Inform	self disclosure	SD
Question	open question	OQ
Question	closed question	CQ
Raise concern		RC
Reflect	Repeat	RE
Reflect	Rephrase	RP
Reflect	Paraphrase	PP
Reflect	Summarize	SU
Reframe		RF
Support		SP
Structure		ST
Warn		WA
Client behaviour counts		
Ask		
Follow		
Resist change		
Change talk		

scales. Some of the second pass therapist behaviour counts can be interpreted as covering those principles, however. "Develop discrepancy" can be seen as represented by asking questions and paraphrasing, "Avoiding discussion" by cooperating with the client, emphasizing the client's responsibility, and personal control in making decisions, "Rolling with resistance" by reframing or paraphrasing the client's statements about change and "Supporting self-efficacy" by affirming and giving expressions of appreciation for the client's efforts towards change.

How well are the traps (Miller and Rollnick, 1991) that should be avoided represented in the MISC? The therapist behaviour count "confront", in the second pass, can be seen as representing the confrontation, labelling, and blaming traps. The remaining traps, however, are not clearly covered in the MISC. According to Miller, the question-answer trap can be detected by tallying the number of open and closed questions and their ratio to reflections. A low score on the first pass therapist rating "egalitarianism" indicates a counsellor had fallen into the expert trap. The premature-focus trap, however, is not at all represented in any of the therapist behaviour counts. In conclusion, it seems that the five general principles of MI are present in the MISC, although somewhat unbalanced, but that not all traps are clearly covered by the MISC. There could be a risk of not assessing all its key elements when applying the MISC as an adherence coding system of MI.

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

Method

Coders

Four advanced students of clinical psychology and a psychologist served as coders in this study. All coders were female. The mean age was around 27 years (range from 22–45 years) All were familiar with MI through reading and training by the first two authors, the second of whom had been trained by Miller and colleagues.

Training time was around 30 hours. All the raters had received the MISC manual before the first session and had studied it in detail. The first session was used to explain and discuss the manual, the global scales and the behaviour counts and their definitions. Eight transcribed prototypical MI sessions (Miller, Rollnick and Moyers, 1998) were used for training in the MISC. Each global rating and each behaviour count was discussed in detail with the first author until the coders appeared to have a good grasp of how to score. The coders were trained until the interrater agreement on the second pass was more than 80% between all coders. The training extended over 2 months and training sessions were held twice a week. To prevent observer drift, all coders met periodically in groups to discuss their ratings.

Interviews

The taped interviews used in this study were recorded MI role-play training sessions. These records were made as part of a seminar on MI organized as in-service training in the Jellinek Centre Amsterdam, a large centre for treating addictions. All counsellors were asked to participate in the study by recording a brief anonymous MI role-play training session. Couples of trainees performed a role-play and were counsellor or client by turns. Most counsellors were social workers (80%) with an average of 5 years of professional experience. Although most were familiar with cognitive behavioural therapy, this workshop was a first training in MI skills.

The role-play instruction was to motivate a simulated client towards behaviour change following the principles and techniques of MI. The two simulated clients had comparable drinking problems. One was a 40-year-old married nurse and the other a 45-year-old male, both drinking too much alcohol, which caused problems both at home and at work. Neither was motivated for change. The role-play was to last a maximum of 10 minutes.

Thirty-nine MI role-play training sessions were understandable and transcribed. The average length of the sessions was 7 minutes and 50 seconds with a range from 4 minutes and 11 seconds to 17 minutes and 30 seconds.

Analysis

Reliability was investigated in different ways. Firstly, the interrater agreement between two coders on the global ratings was calculated using the Gower coefficient (Figure 1). The Gower coefficient computes the absolute agreement between coders for each score given (Gower, 1971) in the range 0 to +1. As the formula in Figure 1 shows, two coders (x and y) are compared on global ratings (i). A score of 1 means that two coders do not differ in their ratings whereas 0 means that they show maximum difference in all their ratings.

Secondly, the interrater reliability for all five coders together on the global ratings was calculated with the intraclass correlation coefficient (ICC). The ICC gives an estimate for the

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

$$1 - \frac{\sum_{i-1}^{n} |x_i - y_i|}{nR}$$

Figure 1. Formula Gower-coefficient

reliabilities of independent raters. A relatively strict ICC, viz the random effects model (Shrout and Fleiss, 1979) [2,1], was used for the 39 MI role-play training sessions coded by the five coders. Both the coders and the interview sessions were regarded as random effects.

Thirdly, to assess the reliability of the second pass codings, the total frequency of classifications of behaviour counts was counted for each coder. Chi-square analyses were performed to examine the similarity between the coders for each second pass behaviour count (Lawlis and Lu, 1972). Chi-square involved a 1×5 (classification by coder) comparison.

Finally, to examine the validity of the MISC, the frequencies of the therapist behaviour counts (TBC) in the observed MI role-play training sessions were compared with the frequencies as they occur in a prototypical MI session. These frequencies in a prototypical MI session come from data obtained by de Jonge, Schaap and Schippers (2000). They transcribed and coded eight MI sessions by acknowledged experts in the field – the tapes illustrating the style – that were provided by the developers (Miller et al., 1998). Three role-plays that displayed the highest resemblance to a prototypical MI session were selected from the present data of 39 MI role-play training sessions. The transcripts of these three MI role-plays were closely examined to determine whether they could, indeed, be evaluated as model MI sessions. In particular, it was judged whether they succeeded in avoiding the traps in MI that Miller and Rollnick (1991, 2002) described.

Results

Table 2 shows the reliability indices for each global rating between the five coders in the 39 MI trainee role-plays. The interrater agreement between two coders, the Gower coefficients (GC, Gower, 1971), on the first pass global therapist ratings range between 0.77 and 0.90, indicating 77 to 90% agreement on absolute scores. The mean GC is above 0.83 for each therapist rating scale. The GCs for the first pass client rating scale range from 0.79 to 0.91, indicating between 79 and 91% absolute agreement. The mean GC is above 0.83 for each client rating scale. The mean GCs for "collaboration" and "benefit" are 0.87 and 0.94, ranging between 0.85 to 0.97.

Table 2 also shows the intraclass correlations (ICC, Shrout and Fleiss, 1979) between the five coders on each of the first pass global rating scales. Cicchetti (1994) formulated guidelines for the interpretation of ICCs. An ICC below 0.40 is poor, between 0.40 to 0.59 it is fair, between 0.60 and 0.74 it is good, and above 0.75 it is excellent. The ICCs for the global therapist ratings are generally poor, ranging from 0.06 to 0.40, with the ICC for genuineness being very poor. The ICCs of the client rating scales are poor to fair, ranging from 0.20 to 0.40. The ICC for "collaboration" is 0.40 and for "benefit" is 0.42, so that both are rated as fair.

Table 3 summarizes the results of the chi-square analyses comparing the total frequency on each count for all coders. The total frequencies for the therapist behaviour counts of affirm, confront, personal feedback, rephrase, paraphrase and support differ significantly between the

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

Table 2. Motivational Interviewing Skill Code: first pass subscale reliabilities based on Gower-coefficients (GC) and intraclass correlation coefficients (ICC)

Global rating	Subscale	GC Mean	GC (range)	ICC
Therapist	Acceptance	.85	(.81–.89)	.24
	Egalitarianism	.83	(.77–.88)	.37
	Empathy	.85	(.81–.87)	.40
	Genuineness	.85	(.79–.89)	.06
	Warmth	.83	(.79–.87)	.36
	Spirit	.85	(.82–.90)	.40
Client	Affect	.84	(.79–.88)	.40
	Cooperation	.86	(.81–.91)	.26
	Disclosure	.87	(.85–.89)	.44
	Engagement	.83	(.79–.86)	.28
Interaction	Collaboration	.87	(.85–.90)	.40
	Benefit	.94	(.91–.97)	.42

Table 3. Second pass behaviour count frequencies by coders

Second pass behaviour count	Rater 01	Rater 02	Rater 03	Rater 04	Rater 05	χ^2 $df = 4$
AD+	6	4	10	4	4	4.40
AD−	12	5	21	12	19	9.84*
AF	*3*	*14*	*3*	*1*	*1*	*24.5***
CO	*6*	*31*	*17*	*10*	*14*	*23.25***
DI	4	2	7	8	3	4.91
EC	8	10	13	7	12	2.2
FA	130	120	164	149	135	8.31
FI	32	23	27	21	19	3.20
GI	48	53	61	53	49	1.71
SD	28	21	40	36	29	7.00
PF	*2*	*25*	*15*	*22*	*18*	*19.19***
OQ	170	155	144	154	137	2.64
CQ	307	289	329	249	326	6.41
RC	6	1	2	8	12	7.30
RE	36	39	35	42	44	0.91
RP	*133*	*171*	*122*	*86*	*127*	*28.75***
PP	*75*	*39*	*90*	*103*	*76*	*29.93**
SU	78	67	63	77	70	2.32
RF	1	5	12	5	16	10.20
SP	*32*	*17*	*37*	*11*	*34*	*17.78***
ST	54	59	50	50	53	1.03
WA	0	0	1	0	1	2.10
Ask	37	23	33	44	32	6.95
Follow	*933*	*709*	*752*	*726*	*706*	*48.23***
Resist change	192	197	194	172	212	4.24
Change talk	46	71	64	75	60	8.02

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

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

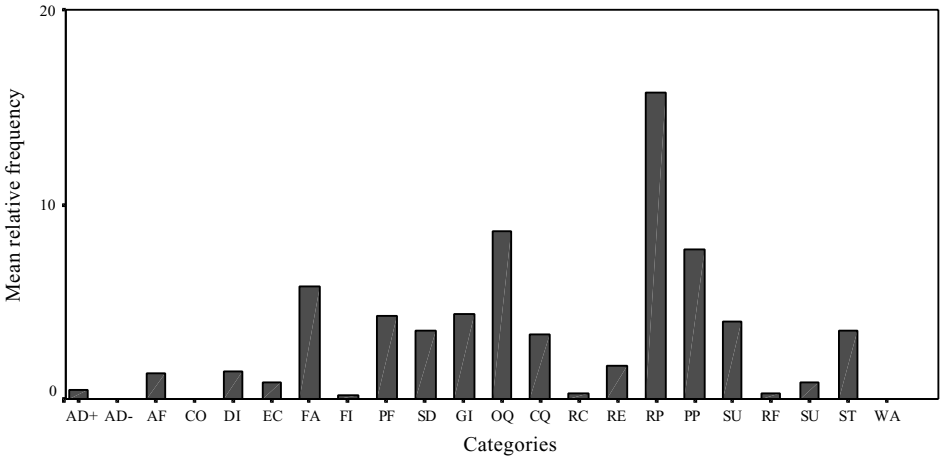

Figure 2. Mean frequencies of therapist utterances in a 7.50 minute prototypical MI session

five coders. Rater 2 differed from the others on the total frequency for affirm and confront. Rater 1 observed fewer expressions of personal feedback and Rater 4 categorized rephrasing and paraphrasing utterances differently. The total frequencies on the client behaviour count differed significantly between the coders. Rater 1 observed more client utterances.

The frequencies of the therapist behaviour counts (TBC) in the observed MI role-play training sessions were compared with the frequencies as they occur in a prototypical MI session in order to examine the discriminant validity of the MISC. Data from de Jonge et al. (2000) were used to obtain a profile of TBC of a prototypical MI session. Eight MI sessions by acknowledged experts in the field (Miller et al., 1998) were transcribed and coded by the MISC. Because the mean length of a MI role-play training session was 7 minutes and 50 second, the first 7 minutes and 50 seconds of the eight expert MI sessions were selected. The mean frequencies on the second pass TBC were calculated. This allowed a profile of a prototypical MI session to be obtained (Figure 2). This profile was compared with the profiles of all the 39 MI training role-plays.

The profiles of the role-plays were obtained by calculating the means, for five coders, on each second pass TBC. Subsequently, the absolute differences in frequencies on the second pass TBC, between the prototypical MI session and each MI role-play, were calculated. These differences were totalled.

The three MI training role-plays with the highest resemblance to a prototypical MI session are shown in Figures 3 to 5. The transcripts of these well-conducted MI training sessions were examined to see whether these counsellors fell into a trap to be avoided in MI.

Examination of these MI training role-plays revealed that, contrary to the prototypical counsellors, these counsellors asked more questions than they rephrased or reflected the client's statements. All counsellors seem to have fallen into the question-answer trap. The following transcript is an example of the question-answer trap. The counsellor in this transcript is the one in the MI role-play profile shown in Figure 3.

T: Tell me, what can I do for you?
C: Do for me, yes I am here, I have problems.

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

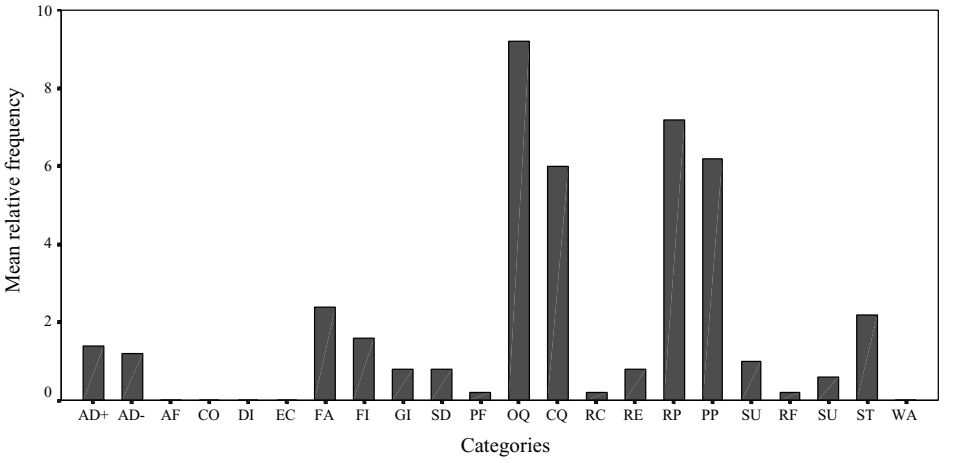

Figure 3. Mean frequencies of therapist utterances in a 7.50 minute MI role-play

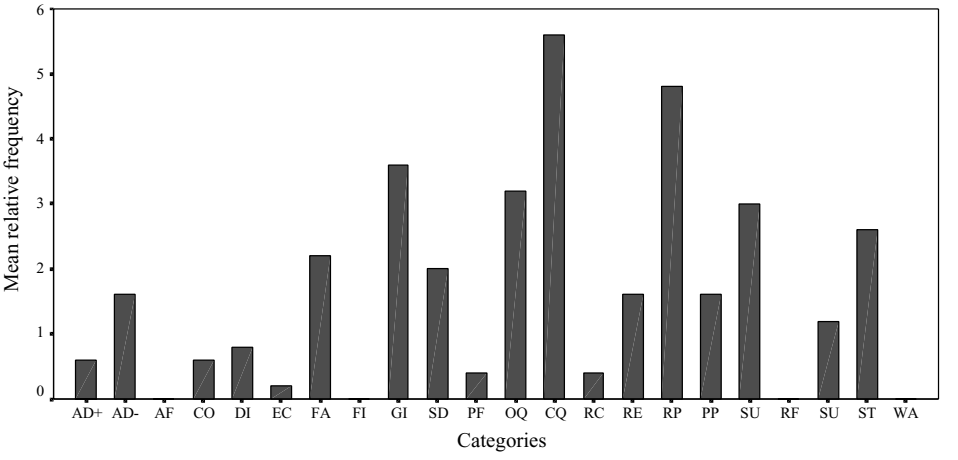

Figure 4. Mean frequencies of therapist utterances in a 7.50 minute MI role-play

T: What kind of problems?
C: Yes a lot of problems, at home and uh drinking. Lately I am drinking more than I used to do.
T: How much more?
C: A lot more. Actually, four or five glasses at nights and uh during the weekends even more.
T: And when did it start?
C: It varies. I used to drink a lot more when I was younger while going out with friends. I used to drink twelve beers or more on an evening. I still like a drink, I used to drink four or five on an evening. But I quit when we got children.

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

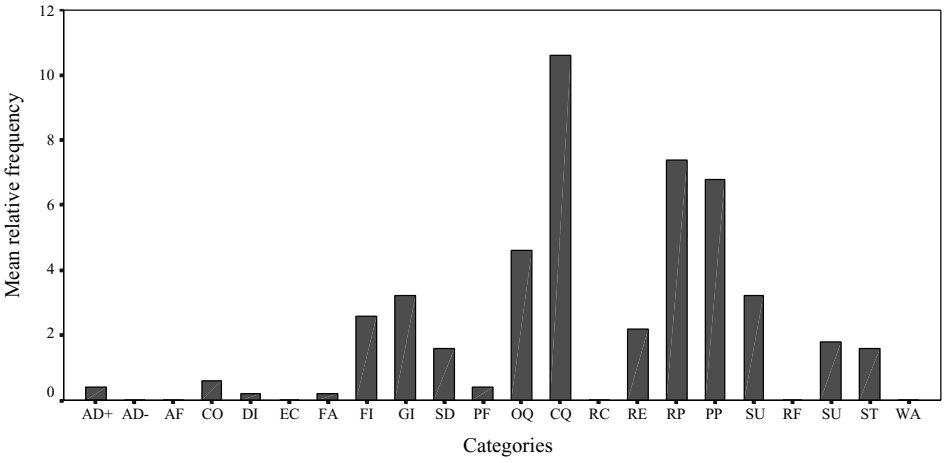

Figure 5. Mean frequencies of therapist utterances in a 7.50 minute MI role-play

T: What did you say? Since the children left the house?
C: No. My children haven't left the house. We have teenagers, difficult teenagers.

Even though the mean score of the five raters on the global therapist behaviour "egalitarianism" is a 4 (range 3–5), indicating that the counsellor's behaviour is neither egalitarian nor authoritative, examination of the transcript of Figure 4 shows that the counsellor falls into the expert trap twice. In talking about the pros and cons of drinking alcohol, the counsellor tells the client what the effects of alcohol are for the client, as the next transcript shows.

T: What negative consequences do you experience while you are drinking? Because there are of course a lot of positive consequences, you will probably be more relaxed (*expert*).
C: Yes, especially after the first two drinks I feel comfortable. I would like to stop after two drinks.
T: That doesn't happen.
C: No it doesn't. If I don't drink I do not know how to relax.
T: That would be one of the consequences. Image yourself 19 years from now, how would your life look like? What do you want in the future?
C: I want to live a happy life and then uh the same life as I live now and then without drinking.
T: Without drinking, but what do you need to do to reach that point? Because according to me, you will have problems if you quit at once (*expert*). You have been drinking for such a long time!

Finally, examination of the transcript of the MI role-play represented in Figure 6 shows that this counsellor falls into the premature-focus trap. During the consultation the client is angry. He wants sleeping pills for his insomnia but has the feeling that everyone wants him to quit

https://doi.org/10.1017/S1352465804001948 Published online by Cambridge University Press

his drinking. The counsellor wants to find out the reasons for the client's drinking problem as the following and last transcript shows.

C: Yes, I have to answer all these questions and I don't want to do that. I just want sleeping pills, what is the problem with that? There are millions of people using them and why can't I get some?
T: No, you don't have to be a criminal for that.
C: No, but it feels like that and the doctor also interrogated me and you do it again. Why bother? Just give me some, I am not asking for your wallet, I just want to sleep and here I apparently don't get them either. I don't understand what I am doing here.
T: We are here to find out some other possible reasons for your drinking problem *(premature-focus*).
C: Yeah, but there is nothing to find out. I mean, I just told you my father died and all these problems with my marriage and everything, my children being irritating and the stress at my work. Everybody is nagging me about my drinking but I don't harm anyone, I just sleep badly.

Discussion

The first aim of the present study was a critical appraisal of the content validity of the MISC. All the general principles of MI (Miller and Rollnick, 1991, 2002) can be identified in the MISC, although they are somewhat unbalanced. The first principle, to *express empathy*, is most elaborately covered by four of the six first pass global rating scales. The other principles, to *develop discrepancy*, *to avoid discussion*, *to roll with resistance* and *to increase self-efficacy* can be found in combinations of second pass therapist behaviour counts. However, the traps to avoid in MI are not so clearly covered by the MISC. Of course, all violations of MI lead to a negative score on the first pass rating scale *spirit of MI*. However, when using the MISC in training, or in testing the integrity of treatment, a more complete and clear measurement is desirable. The MISC is not as optimal as desired for training and treatment integrity.

Furthermore, just one second pass therapist behaviour count, *confront*, can be seen as representing three MI traps, namely *confront-denial*, *blaming* and *labelling*. The *question-answer* trap can be discovered by comparing the second pass questions and the reflections ratio. Tallying the questions and reflections and calculating the ratio is rather laborious, however. The *expert* trap can be seen as a low score on the "egalitarianism" scale. Finally, the *premature-focus* trap does not seem to be explicitly described in any of the first or second pass therapist behaviours. The MISC does not appear to cover all aspects of MI. One could pose the question whether that restricts the usage of the MISC as an adherence instrument. Investigation of the discriminant validity of the MISC, by comparing a prototypical MI session with the MI role-play training sessions, shows that even when sessions show a high resemblance to prototypical MI, this does not signify the absence of the traps to avoid in MI.

The second aim of the present study was an investigation of the reliability of the MISC as used by five coders coding 39 MI role-play training sessions of different counsellors. As the high Gower coefficients show, the coders agreed closely on the absolute scores but the interrater reliabilities between five coders estimated by the intraclass correlations were mainly poor. This could be caused by restriction of range, as the MI training role plays resembled each other. Where Tappin et al. (2000) expected that more raters and more counsellors might

improve the reliability of the MISC, our data did not support that expectation. Moyers, Martin, Catly, Harris and Asluwalia (2003) also studied the reliability of the MISC. They found higher ICC scores for the therapist ratings for two coders coding 86 MI session than were found in this study, but the reported ICCs were poor for the client rating effect of disclosure and fair for the interaction rating collaboration. So it seems difficult, despite intensive pre- and ongoing training, reliably to code the client and interaction rating scales.

The interrater reliability of the second pass behaviour counts is reasonable. Although the coders differed significantly on 8 of the 24 behaviour counts (30%), the differences can be explained. Coder 2 coded more AF and less SP and perhaps swapped these therapist behaviour counts. Coder 2 coded more CO than the others, resulting in a lower level of agreement between all coders. CO, however, was rarely present in the 39 MI role-play training sessions and this might have influenced the interrater agreement. Coder 4 confused the classifications of PP and RP, which is very understandable because both are subclassifications of the therapist behaviour count *reflect*. Moyers et al. (2003) found poor and fair agreement on 30% of the subclassifications as well and on the same therapist behaviour counts, viz the affirm, confront and reflect categories. They also stated that the rarely displayed therapist utterances such as AF and CO are hard to code reliably.

The total frequency on the second pass client behaviour count did not differ between the raters, except for the count *follow*. Coder 1 coded more neutral client utterances. However, this does not influence the impression of the client motivation for change because this depends on the number of *change talk* or *resistance* utterances. More explicit descriptions of the total behaviour counts, however, can improve correct classifications and increase the interrater agreement.

The investigation of the reliability of the MISC in this study can be criticized. Firstly, there might have been several sources of coder bias. All coders were female. This might have caused coder bias, even though findings in other rating studies are not that clear. Gale, Woodward, Hayes, Sivakumaran and Hanssen (2002) found that women were more cautious then men in their ratings but Hill, O'Grady and Price (1988) did not find gender differences when studying rater bias. A second source of coder bias was the training of the coders. The first author trained the others in the use of MISC, but had not been trained in the MISC by the original authors. The training and coding process relied heavily on the MISC coding manual and the coders' interpretation. Although coding manuals should be self-guiding and should not be dependent upon the originating authors, our interpretation might have resulted in some biases.

A second limitation was the use of MI role-play training sessions in this study. These were short role-plays in which fellow trainee-counsellors performed as counsellors and as clients. It is possible that this influenced the variety of therapist behaviours that could be observed and might have influenced the interrater agreement. Although the study of longer and more naturalistic interaction would enrich the data, this appears not to be essential for investigation of the MISC.

A suggestion for future studies of the MISC reliability, therefore, is assessment of the interrater agreement between the coders in longer consultations. Longer MI sessions give more coding material and will make judging therapist behaviour easier. In conclusion, the MISC is a useful instrument to use during process research, for studying MI in depth and for identifying its active agents. Process research needs detailed research tools to examine what happens in treatment sessions. An additional goal is to link process data to treatment outcome (Hill and Lambert, 2004). However, the representation of the general MI principles

is somewhat unbalanced and incomplete as far as the traps are concerned. This perhaps makes the MISC less useful for documenting a counsellor's adherence to the clinical style of MI. What is needed is an instrument coding the MI style and the traps to avoid in MI. Furthermore, it is also possible to evaluate an MI session in a different way, viz on a content level. A therapist prepares a client to make a well-considered choice about change and this should be done by following a particular sequence of questions or utterances each session. To use the MISC during MI training of counsellors may be too labour-intensive because it takes three passes to code a MI session with the MISC. So a more practical and complete instrument (in particular, indeed, for non-expert helpers) is needed. Maybe the Coding System for Integrity of Treatment – Motivational Interviewing, to assess the MI style, traps and the content of MI that we are currently in the process of developing can play a role in this respect.

Acknowledgements

The authors are grateful to Dr W. R. Miller for his comments on the original draft and his reaction on our appraisal of the MISC.

References

Chamberlain, P., Patterson, G., Reid, J., Kavanagh, K. and Forgatch, M. (1984). Observation of client resistance. *Behavior Therapy, 15*, 144–155.

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

De Jonge, J. M., Schaap, C. P. D. R. and Schippers, G. M. (2000). *Therapeutic Interventions that Trigger Self-motivational Statements: a sequential analysis of prototypical motivational interview sessions*. Poster presented at the International Conference of Addictive Behaviors, Cape-Town.

DiClemente, C. C. (1991). Motivational interviewing and the stages of change. In W. R. Miller and S. Rollnick (Eds), *Motivational Interviewing Preparing People to Change Addictive Behavior* (pp. 191–202). New York: The Guilford Press.

Dunn, C., Deroo, L. and Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. *Addiction*, *96*, 1725–1742.

Gale, T. M., Woodward, A., Hawley, C. J., Hayes, J., Sivakumaran, T. and Hanssen, G. (2002). Risk assessment for people with mental health problems: a pilot study of reliability in working practice. *International Journal of Psychiatry in Clinical Practice*, *6*, 73–82.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27*, 857–874.

Hill, C. E. and Corbett, M. M. (1993). A perspective on the history of process and outcome research in counseling psychotherapy. *Journal of Counseling Psychology*, *40*, 3–24.

Hill, C. E. and Lambert, M. J. (2004). Methodological issues in studying psychotherapy process and outcomes. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 84–135). New York: John Wiley and Sons.

Hill, C. E., O'Grady, K. E. and Price, P. (1988). A method for investing sources of rater bias. *Journal of Counseling Psychology*, *35*, 346–350.

Kendall, P., Holmbeck, G. and Verduin, T. (2004). Methodology, design, and evaluation in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 16–43). New York: John Wiley and Sons.

Lawlis, G. F. and Lu, E. (1972). Judgement of counseling process: reliability, agreement, and error. *Psychological Bulletin*, *78*, 17–20.

Miller, W. R. (1983). Motivational interviewing with problem drinkers. *Behavioural Psychotherapy*, *11*, 147–72.

Miller, W. R. (2000). *Motivational Interviewing Skill Code*. Unpublished Document.

Miller, W. R. and 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. and Rollnick, S. (1991). *Motivational Interviewing: preparing people to change addictive behavior*. New York: The Guilford Press.

Miller, W. R. and Rollnick, S. (2002). *Motivational Interviewing: preparing people for change*. New York: The Guilford Press.

Miller, W. R., Rollnick, S. and Moyers, T. B. (1998). Motivational Interviewing, professional training videotape series. University of New Mexico, Albuquerque, USA.

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

Noonan, W. C. and Moyers, T. B. (1997). Motivational interviewing. *Journal of Substance Misuse*, *2*, 8–16.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–429.

Tappin, D. M., McKay, C., McIntyre, D., Gilmour, W. H., Cowan, S., Crawford, F., Currie, F. and Lumsden, M. A. (2000). A practical instrument to document the process of motivational interviewing. *Behavioural and Cognitive Psychotherapy, 28*, 17–32.