Computers in Human Behavior 44 (2015) 284–292

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Coding interactions in Motivational Interviewing with computer-software: What are the advantages for process researchers?

Florian E. Klonek [a,*], Vicenç Quera [b], Simone Kauffeld [a]

[a] Technische Universität Braunschweig, Institute of Psychology, Braunschweig, Lower Saxony 38106, Germany
[b] Universitat de Barcelona, Facultat de Psicologia, Dept. of Behavioral Science Methods (Dept. de Metodologia de les Ciències del Comportament), Spain

ARTICLE INFO

Article history:

Keywords:
Motivational Interviewing
Motivational Interviewing Treatment
Integrity code
Software coding
Observational methods
Timed-event data
Behavioral slicing

ABSTRACT

Motivational Interviewing (MI) is an evidence-based behavior change intervention. The interactional change processes that make MI effective have been increasingly studied using observational coding schemes. We introduce an implementation of a software-supported MI coding scheme—the Motivational Interviewing Treatment Integrity code (MITI)—and discuss advantages for process researchers. Furthermore, we compared reliability of the software version with prior results of the paper version. A sample of 14 double-coded dyadic interactions showed good to excellent interrater reliabilities. We selected a second sample of 22 sessions to obtain convergent validity results of the software version: substantial correlations were obtained between the software instrument and the Rating Scales for the Assessment of Empathic Communication. Finally, we demonstrate how the software version can be used to test whether single code frequencies obtained by using intervals shorter than 20 min (i.e., 5 or 10 min) are accurate estimates of the respective code frequencies for the entire session (i.e., behavior slicing). Our results revealed that coding only a 10-min interval provides accurate estimates of the entire session. Our study demonstrates that the software implementation of the MITI is a reliable and valid instrument. We discuss advantages of the software version for process research in MI.

1. Introduction

Motivational Interviewing (MI) is a collaborative, client-centered communication style that aims to strengthen intrinsic motivation for change. As communication is at the core of an MI intervention, there have been an increasing amount of process studies that have used systematic observational methods to assess verbal behavior in MI—either focusing on therapist behaviors (e.g., D'Amico et al., 2012; Madson, Campbell, Barrett, Brondino, & Melchert, 2005), client language (e.g., Baer et al., 2008; Hodgins, Ching, & McEwen, 2009; Houck, Moyers, & Tesche, 2013; Martin, Christopher, Houck, & Moyers, 2011), or both (e.g., Hannöver, Blaut, Kniehase, Martin, & Hannich, 2013; Vader, Walters, Prabhu, Houck, & Field, 2010). Traditionally, process studies have used paper–pencil methods to assess verbal behaviors in MI (e.g., Baer et al., 2008; D'Amico et al., 2012). However, if researchers also want to capture dynamic and sequential information about dyadic interaction in MI, they also need to record on- and offset times of participants' verbal behavior. As tallying behaviors and recording on- and offset times of verbal interactions requires many attentional resources (Bakeman & Quera, 2011), we propose that researchers should rely on computer-supported coding instruments. Computer-supported coding instruments can automatically record time and sequential information whereas observers can focus only on coding the functional aspects of the interaction.

1.1. Comparing paper-and-pencil and computer-assisted administration of observational measures

There are many studies that have compared computer-assisted vs. paper–pencil administration of interventions (Butler & Correia, 2009; Murphy, Dennhardt, Skidmore, Martens, & McDevitt-Murphy, 2010; Serowik, Ablondi, Black, & Rosen, 2014) or questionnaires (e.g., Booth-Kewley, Larson, & Miyoshi, 2007; Feigelson & Dwight, 2000; Finegan & Allen, 1994; Rosenfeld et al., 1991; Rammstedt, Holzinger, & Rammsayer, 2004; Sharp & Hargrove, 2004). The underlying assumption of these studies is that the characteristics of an instrument may change when it is transferred from the paper–pencil version to a computer-based procedure.

* Corresponding author at: Technical University Braunschweig (Technische Universität Braunschweig), Dept. of Industrial/Organizational and Social Psychology, Spielmannstr. 19, 38106 Braunschweig, Germany. Tel.: +49 (0)531 391 2566; fax: +49 (0)531 391 8173.

E-mail address: f.klonek@tu-bs.de (F.E. Klonek).

http://dx.doi.org/10.1016/j.chb.2014.10.034
0747-5632/

Table 1
Example of a two event sequence coded by two observers with perfect agreement for summary scores but zero agreement for point-by-point agreement.

	Behavioral event n_i	Behavioral event n_{i+1}	Summary score
Transcript	"You are unsatisfied with the current situation?" (Closed Question)	"You feel that things should change." (Simple Reflection)	
Observer 1	Code A	Code B	1 Code A 1 Code B
Observer 2	Code B	Code A	1 Code A 1 Code B
Point-by-point agreement	No agreement	No agreement	

Some researchers have argued that computer support may assist in the reduction of cognitive load (Barzilai & Zohar, 2006; Salomon, Perkins, & Globerson, 1991). Using the computer as a "cognitive tool" (Jonassen, 1995) that "reduces the load of lower order thinking processes, such as memory or computation" (Barzilai & Zohar, 2006, p. 141) gives observers more cognitive resources for the actual coding task. In the field of observational research, researchers often state that the process of systematic coding is labor intensive and demands a high amount of cognitive workload (Bakeman & Quera, 2011; Baumeister, Vohs, & Funder, 2007; Meinecke & Lehmann-Willenbrock, in press). Observers have to parse the interactional dynamics into different observational units, keep in mind the codes of the observational scheme, and consequently assign them to the observational unit. Conversely, there is an argument that a retrieval of information on computer-screens is tiring and may increase cognitive demands (Wästlund, Reinikka, Norlander, & Archer, 2005). As a result, a computer-supported observational instrument could either enhance coding work by "off-loading" cognitive demands (Barzilai & Zohar, 2006), or impede coding work by mobilizing "perceptual and executive cognitive resources" (Wästlund et al., 2005, p. 390). As methodological studies in observational research have shown that an increase in cognitive load increases observational error rates (La France, Heisel, & Beatty, 2004, 2007), it is important to investigate whether computer-supported and paper–pencil instruments actually provide equivalent results. Nonetheless, the equivalence between computer-assisted vs. paper–pencil administered coding of *observational instruments* has seldomly been addressed in empirical studies. The present study seeks to address this gap by focusing on a paper–pencil coding instrument that is extensively used in research and practical evaluations of MI (e.g., Madson et al., 2005; Moyers, Martin, Catley, Harris, & Ahluwalia, 2003).

With respect to economy, research has suggested that computer-assisted coding can save time and money in a research project (Glynn, Hallgren, Houck, & Moyers, 2012; Tapp et al., 2006). Tapp et al. (2006) reported that a software version of the Observational Ratings of Caregiver Environment (M-ORCE; Gunnar, Kryzer, Phillips, & Vandell, 2001) was more time efficient and more accurate in comparison to a paper–pencil version of this instrument. In addition, the computer version saved an average of 1 h and 13 min per subject per observation unit. The authors calculated that a project with 20 subjects and 20 observation units might save $6800 in personal costs (based on an average wage of $14 per hour)—an amount of money that easily outweighs the initial investment in software and computer equipment. Similarly, Glynn et al. (2012) developed free software for coding of verbal interactions in MI using WAV-format audio files. The authors reported that computer software saved at least 7500 work hours and $60,000 in comparison to the use of transcript-based methods.

Computer-assisted coding also facilitates obtaining timed-event data (Bakeman & Quera, 2011) because observers who use paper–pencil assessments would need to keep track of time while simultaneously coding behavior, a task that demands additional cognitive resources and might result in less accurate coding. By contrast, computer administered coding permits recording time and sequential information automatically. In addition to this, computer-assisted observational instruments allow coders to recode more easily in comparison to having to rewind and play a tape by hand (e.g., INTERACT from Mangold, 2010 or Noldus Observer XT, e.g., Clayman, Makoul, Harper, Koby, & Williams, 2012).

1.2. Importance of assessing point-by-point agreement in motivational interviewing

As computer-assisted coding systems make it possible to obtain timed-event data (i.e., sequences of behavioral codes for which onset and offset times have been recorded; see Bakeman & Quera, 2011), they also allow investigators to find out whether observers agree at a more fine-grained level (Yoder & Symons, 2010). Previous psychometric studies in MI research usually have reported Intraclass Correlation Coefficients (ICCs; Fleiss & Shrout, 1978) as reliability measures (Brueck et al., 2009; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005). Using ICC as a reliability estimate offers a different approach compared to using a point-by-point agreement index, such as Cohen's (1960) kappa. ICCs only take into account whether observers were able to agree about summary, or global, behavior measures (frequencies and durations) per observation unit. By contrast, kappa estimates their point-by-point agreement and therefore constitutes a more conservative approach (Bakeman & Quera, 2011; Yoder & Symons, 2010). Table 1 gives an example that illustrates this crucial difference. The two observers from the example in Table 1 agree perfectly to the overall sum for both codes A and B, whereas point-by-point agreement is zero.

In the case of MI, scholars have stressed that interviewers should formulate reflections instead of closed questions. In many cases, reflections are nearly identical to closed questions, differing only in terms of the voice intonation at the end of a sentence (cf., Table 2). The number of reflections is considered to be a competence indicator in the assessment of MI competency (Moyers, Martin, Manuel, & Miller, 2003). In addition, coding a closed question instead of a reflection will additionally reduce the interviewers' competency level in terms of relative amount of open questions to all questions. As a result, confusing a reflection with a closed question will have a twofold negative effect on the treatment adherence of an interviewer; therefore, establishing point-by-point agreement is vital in that case. The software-supported version solves this because it saves sequential and onset/offset times of coded behaviors automatically; that is, without the need of human observers to additionally note the times during their coding work.

1.3. Preserving time information with software-support

Moyers et al. (2005) developed the MI Treatment Integrity code (MITI). It is a paper–pencil administered observational instrument that can be applied by trained observers to count behaviors that are considered to be good indicators of MI: Open Questions, Simple

Table 2
Verbal codes used for the observational instrument (MITI).

Code	Verbal example
Giving Information *The interviewer gives information, educates, provides feedback from assessment instruments, discloses personal information or gives an opinion without advising*	"Energy-saving light bulbs are an ecological alternative and can save up to 80% of electricity" "There are special depositary available where materials with heavy metals can be recycled"
Open Question *Is used for questions that allows a wide range of possible answers. It invites the conversational partners' perspective or may encourage self-exploration*	"What are the advantages of changing this behavior?"
Closed Question *Is used for questions that can be answered with a "yes/no" response or if the question specifies a very restricted range of answers*	"Do you think this is an advantage?" "You are not sure what is going to come out of this talk?" (the question mark and voice intonation at the end marks the difference to a reflection)
Simple Reflection *Repetition, rephrase, or paraphrase of a conversational partners' previous statement*	A: "I do not believe this conversation will change anything" B: "You are not sure what is going to come out of this talk"
Complex Reflection *Repeats or rephrases the previous statement of a conversational partner but adds substantial meaning to it*	A: "Mostly, I would change for future generations – for their advantage. If we waste everything, then there will be nothing left" B: "It sounds like you have a strong feeling of responsibility"
MI Adherent *Captures behaviors that are consistent with a motivational interviewing approach, e.g., asking permission before giving advice, affirming, emphasizing the conversational partners control and autonomy, supporting him/her with compassion and sympathy*	"You should be proud of yourself for your past's efforts" (affirming) "I'm here to help you with this" (compassion) "If you agree with it, we could try to brainstorm some ideas that might help you" (asking permission before giving advice)
MI Non-Adherent *Captures behaviors that are inconsistent with a motivational interviewing approach, e.g., advising without permission, confronting, orders, commands or imperatives*	"So that's really no excuse for wasting energy" (confront) "You should simply scribble a note that reminds you to turn the computer off during breaks" (advise without permission)

Note: **Bold** = Codes used in the MITI.

and Complex Reflections, and MI Adherent Behavior (asking permission before giving advice, affirmations, emphasizing control, support). The MITI also includes behaviors that are considered to indicate interviewers' deficits in MI (MI Non-Adherent behaviors and Closed questions) and neutral behaviors (Giving Information). The authors noted that it did not provide information about how well interviewers adhered to MI in a particular coded segment. Verbatim transcripts of sessions can be used to preserve sequential information of the interaction (Hannöver et al., 2013; Hodgins et al., 2009). Unfortunately, it is very time consuming to prepare transcripts—depending on the quality of the session, the time to create transcripts may take a 20–90 multiple of the total session duration (Graf, Aksu, & Rettinger, 2010; Lalouschek & Menz, 2002). In other words, the additional amount of time to create transcripts for a dataset of 30 sessions lasting, on average, 30 min varies between 300 and 1350 working hours. In the MI process study from Hannöver et al. (2013), the additional costs to create verbatim transcripts added up to 12,000 € (about $16,500 US) or about $100 US per session (W. Hannöver, personal communication, March 5, 2014).

The software version solves this drawback because it preserves time *and* sequential information without the necessity to create transcripts. It allows one to extract MI adherence across specific segments of the observation session and to check for treatment adherence. Moreover, the software version saves the links between coded segments and the video recording, a feature that supervisors may use to qualitatively explore why interviewers performed under or above competency level (Mangold, 2010). Another important feature of the software version is that timed-event data can be used for sequential analyses, which provide information about the behavioral temporal dynamics (e.g., Moyers & Martin, 2006; Moyers et al., 2003).

1.4. Thin behavior slicing: what is the minimal coding interval

The MITI was also developed as a condensed and *economical* observation instrument that (a) saves time because only a random sample of 20 min of the entire MI session is coded, and (b) only includes the seven most important behaviors (Giving Information, Closed and Open Questions, Simple and Complex Reflection, MI Adherent and Non-Adherent behavior; Moyers et al., 2005). Nevertheless, some observational studies have coded MI sessions that were shorter than 20 min (e.g., Gaume, Gmel, Faouzi, & Daeppen, 2008). Using shorter intervals is necessary if the intervention itself lasts less than 20 min, which may be the case, for example, if MI is integrated in a Brief Alcohol Intervention (BAI; Daeppen, Bertholet, Gmel, & Gaume, 2007; Gaume et al., 2008) or Brief Motivational Intervention (BMI; e.g., Peterson, Baer, Wells, Ginzler, & Garrett, 2006). Those studies assumed that the entire range of MI specific behaviors can be accurately observed within that time frame. As these assumptions have not yet been tested, we used the software version to investigate the *minimal coding interval* that is necessary to asses MI specific behaviors (as defined by the MITI). We want to test whether minimal intervals of 5 and 10 min can yield an accurate estimate of the entire session. This procedure can be referred to as *thin slicing* (Ambady & Rosenthal, 1992; Jung, Chong, & Leifer, 2012; Waller, Sohrab, & Ma, 2013).

1.5. Aim and contributions

Our goal was to implement a software version of the MITI and compare its performance with prior results of the paper–pencil German version. Our contributions can be summarized as follows: (a) we performed analyses of reliability and validity in order to know whether the software version of the MITI is equivalent to its paper–pencil counterpart, (b) we obtained point-by-point measures of observer agreement (time-unit kappa indices) for every MITI code, and (c) we tested whether behavioral measures obtained by coding intervals shorter than 20 min (5, 10 min) were accurate estimates of measures for the entire session (minimal coding interval).

2. Method

2.1. Source of data

The data used for this study were obtained from Project EnEff Campus, a research project funded by the German Ministry of Economics and Technology (BMWi). This project investigates how communication skills in MI facilitate participants' engagement in pro-environmental behavior. All procedures of the study were approved by the Institutional Review Boards on data security, and all participants provided written informed consent to be videotaped. A total of 28 interviewers and 77 clients participated in the study. Interviewers were 50% female, with a mean age of 29 years. The majority of interviewers studied psychology ($n = 19$), whereas the remaining indicated a different academic field of study ($n = 9$). Participants (i.e., clients) were 78% female, with a mean age of 24.3 years. A subgroup of fifteen interviewers received training in MI, whereas thirteen interviewers had no previous training in MI. The training in MI was designed according to the eight stages of learning MI (Miller & Moyers, 2006). It contained exercises from the Motivational Interviewing Network of Trainers (2008) manual and encompassed about 40 h within a period of 3 months. Interviewers in the training group practiced MI inter-individually in peer groups on a biweekly basis, received homework, and studied MI literature. Trained and untrained interviewers conducted between 1 and 3 sessions with a unique client, which resulted in a final data pool of 74 recorded sessions. In order to allow independent measurements, interviewers only contributed one interview to subsequent analyses.

2.2. Measures: software-supported MITI-d

We used the German version of the MITI (Brueck et al., 2009; MITI-d). It includes seven interviewer behaviors that are coded by an external observer. The behaviors include: (a) Giving Information, (b) Questions (Open and Closed), (c) Reflections (Simple and Complex), and (d) MI Adherent and Non-Adherent behaviors. Table 2 gives an overview of behavioral codes used in the MITI, their definitions, and verbal examples. The full English manual of the MITI is freely accessible online via http://casaa.unm.edu/download/miti.pdf. Furthermore, three coded transcripts in the online material demonstrate how verbal behavior is coded using the MITI (Energy Manager A–C). In the paper–pencil version of the MITI-d, coders have to count instances of these behaviors and mark a tally on an observational sheet.

The software version of the MITI-d was implemented in the INTERACT 9 videoanalysis program (Mangold, 2010). INTERACT is a commercial video coding and analysis software for observational studies that can be used as an interface for any coding scheme. We have prepared a demonstration video for readers who are interested in the technical implementation of the software MITI (see online material "Videodemo_sMITI.wmv").

2.3. Training of observers

Prior to coding, two student observers received training in the application of the MITI-d. Both observers proceeded through a series of graded learning tasks using fictional transcripts designed by coding experts in MI (Brueck, Frick, & Loessl, 2006; Demmel & Peltenburg, 2006; Project MILES, 2011). The initial phase of training included reading selected chapters from Miller and Rollnick (2002) and the MITI manual (Brueck et al., 2006); also, observers worked with pre-coded transcripts from the author of the German MITI (R. Brueck, personal communication, September 19, 2011) and MI video recordings (Demmel & Peltenburg, 2006) for learning to code MI relevant behavior. Training was also covered by means of further gold standard transcripts and audio material by a second German coding group (Project MILES, 2011). In the practice phase, the two observers were required to code both training transcripts and tapes that had been previously double-coded by each other and the coding coordinator (F.K.). The coded session was jointly reviewed, utterance by utterance, and cases of disagreements were solved by discussing any discrepancies (Yoder & Symons, 2010). The observers and the coordinator had regular meetings to discuss coding problems and other areas of difficulty. None of the samples in the training dataset were used for subsequent analyses in this study.

We also prepared online demonstration material (in German and English) that we coded with the software instrument. This supplemental online material can be used for training and learning purposes by other research groups who are interested in using the software version.

2.4. Training in software use

Both observers also received training from the first author (F.K.) for using the computer-assisted version of the MITI-d, consisting of an introduction to the INTERACT software and an explanation of how to use media and software files in INTERACT. Furthermore, they received guidance on how to use keyboard shortcuts for controlling the latter instead of using the mouse and a search-and-click modus. This supplemental software training took about 2–4 h and was added in order to facilitate coding and economize time resources. Support for software use was given until the observers were comfortable enough to work independently with the software.

2.5. Rating scales for the assessment of empathic communication (REM)

We used the Rating Scales for the Assessment of Empathic Communication in medical interviews (REM, Nicolai, Demmel, & Hagen, 2007) in order to establish convergent validity measures. The REM is a 9-item instrument with 6 items measuring *empathy* and 3 items measuring *confrontation*. Empathy in REM is defined as the interviewer's cognitive ability to perceive and understand the client's perspective and his or her behavioral ability to communicate this understanding (Nicolai et al., 2007). Prior studies have shown acceptable reliability for the REM and convergent validity between the MITI and the REM. We reformulated the terms *doctor* and *patient* from the original version into the more neutral terms *interviewer* and *client* because we did not use medical interviews. The observers rated the extent to which interviewers displayed a specific behavior on a five-point Likert-type scale for a subsample of 22 independent sessions each. The lower and upper ends of the scale are anchored in behavioral descriptions; for example, *the interviewer showed a lot of understanding for the client's point of view* (5 points) vs. *the interviewer showed no understanding of the client's point of view* (1 point).

The observers first listened without pause to the interview and answered the REM for each session. Empathy and confrontation subscales showed good internal consistency for both raters ($\alpha_{(\text{rater1})} = .91$ and $\alpha_{(\text{rater2})} = .91$ for the former; $\alpha_{(\text{rater1})} = .95$ and $\alpha_{(\text{rater2})} = .89$ for the latter).

2.6. Overview of statistical plan

Reliability of the software version of the MITI-d was established by calculating ICCs for fourteen double-coded sessions. The ICC is a statistical index commonly used to estimate reliability because it adjusts for chance agreement and systematic differences between observers (Fleiss & Shrout, 1978; McGraw & Wong, 1996, p. 35); it is therefore a more conservative estimate than the Pearson product moment correlation. We computed absolute (criterion-referenced) agreement between two observers according to the following formula (Bakeman & Quera, 2011), which is stricter than the relative (norm-referenced) agreement (Bakeman & Quera, 2011; McGraw & Wong, 1996): $ICC = [MS_b - MS_e]/[MS_b + MS_e - (2/n)(MS_o - MS_e)]$, where n is the number of observations, or summary values, obtained by each observer; MS_b is the between measures mean square; MS_o is the observer, or repeated measures, mean square; and MS_e is the error mean square.

We classified ICCs according to Cicchetti's (1994) proposed cutoff criteria: below .40 = poor; .40–.59 = fair; .60–.74 = good; and .75–1.00 = excellent. We compared ICCs with a prior reliability analyses from Brueck et al. (2009), who applied the paper–pencil MITI-d in a German sample.

Cohen's kappa coefficients for individual behavioral codes (Bakeman & Quera, 2011; Cohen, 1960) were computed in order to assess point-by-point agreement between observers, according to the following formula: $\kappa = (P_o - P_c)/(1 - P_c)$, where P_o denotes the observed percentage of agreement and P_c denotes the expected percentage of agreement due to chance. Kappa estimates the observed level of agreement between two coders for a set of nominal ratings and corrects for agreement that would be expected by chance. Classical Cohen's kappa assumes that a common reference frame (either time of event units) for the two observers is established such that observers make decisions about how to code at every element or unit, and, consequently, that for every element either an agreement or a disagreement can be tallied. Our observers first independently parsed the stream of behavior into behavior units (i.e., recording their onset and offset times) and then coded them. As we cannot assume that they made coding decisions every time unit (second), classical Cohen's kappa cannot be computed. Instead, we can assume that they made decisions when behaviors started and ended. In that case, the common reference frame must be estimated by a dynamic programming algorithm that provides the optimum alignment between the two observer sequences. Two alternatives for aligning them exist: taking into account either the sequence of behavioral events plus their onset and offset times; or taking into account the event sequence itself without any time information, providing time-based and event-based alignment kappa coefficients, respectively. We used the dynamic-programming event-alignment algorithm from program GSEQ that determines the optimal global alignment between two event sequences and provides such coefficients (Bakeman & Quera, 2011; Bakeman, Quera, & Gnisci, 2009; Quera, Bakeman, & Gnisci, 2007).

Convergent validity was established by calculating Pearson correlations between MITI-d codes and the Rating scales for Empathic Communication. We compared these correlations with a convergent validity of the paper–pencil version from Nicolai et al. (2007).

Two different interval lengths (of 5 and 10 min) within the first 20 min of the interaction were selected as behavior slices. In the first case, the 20 min were divided into 4 segments, or slices, of 5 min each, whereas in the second case they were divided into 2 segments of 10 min each. In order to evaluate whether they provide a good estimate for the entire session, we computed ICCs between code frequencies for the slices and the summary scores for the entire session. We used Cicchetti's cutoff criteria: ICCs above .60 indicated that a slice provided a good estimate of the entire session, whereas ICCs smaller than .60 and above .40 indicated that it provided a fair estimate.

3. Results

3.1. Reliability analysis

Sample size for reliability analysis was chosen a priori following two guidelines: Bakeman, Deckner, and Quera (2005) recommended sampling between 15% and 20% of a corpus to check reliability using the kappa coefficient. Furthermore, the minimum sample size for calculation of ICCs lies around five (Yoder & Symons, 2010), while ten or more sessions will result in more robust results (Bakeman & Quera, 2011). In order to fall above this recommended criteria, we randomly selected a subsample of 14 independent sessions (50% of interviewers) for reliability analyses. Random samples of 20 min were chosen for each session.

Results of observer agreements are shown in Table 3. We compared ICCs calculated for the software MITI-d with ICCs calculated for the paper–pencil version reported by Brueck et al. (2009).

Both observers achieved good to excellent levels of interrater agreement for the summary scores (ICCs from .61 for MI Non-Adherent to .98 for Giving Information). Most ICCs for the software MITI-d were higher than those reported by the paper–pencil version [Giving Information (.98 vs. .62), Closed Questions (.94 vs. .92), Open Questions (.91 vs. 80), Complex Reflections (.91 vs. .53), MI Adherent (.72 vs. .70), and MI Non-Adherent (.61 vs. .01)]. In sum, the ICCs are favorably comparable with the results obtained from the paper–pencil version. Only the code Simple Reflection achieved a marginally higher reliability score in the paper–pencil version (.78 vs. .86). As the 95% confidence intervals for Simple Reflection ICCs overlap (.73–.93 for the paper version vs. .44–.92 for the software version), we can assume that the reported point values for our sample do not differ in the population.

Time-based and event-based kappa coefficients, as recommended by Bakeman et al. (2009) and computed by the GSEQ software (Bakeman & Quera, 2011), are shown in Table 3. Codes were matched utterance-by-utterance using the alignment algorithm implemented in GSEQ (Bakeman et al., 2009). Whereas time-unit kappas tend to overestimate the true agreement, event-based kappas tend to underestimate it. Hence "their range likely captures the 'true' value of k[appa]" (Bakeman et al., 2009, p. 146). Hence, the average of time-based and event-based kappas for each code are categorized in Table 3 according to Sachs' (1999) cutoff criteria:

Table 3
Summary (ICCs) and point-by-point agreement (kappas) for the paper–pencil and the software version.

Behavioral code	Paper-and-pencil MITI-d[a]	Software MITI-d				
	Summary scores	Summary scores		Point-by-point agreement		
	ICCs (CI)	ICCs (CI)	Category[b]	Kappa (TU)	Kappa (E)	Category[c]
Giving Information	.62 (.11–.84)	.98** (.93–.92)	Excellent	.83	.75	Strong
Closed Questions	.92 (.79–.96)	.94** (.82–.98)	Excellent	.82	.70	Strong
Open Questions	.80 (.61–.90)	.91** (.74–.97)	Excellent	.80	.80	Strong
Simple Reflections	.86 (.73–.93)	.78** (.44–.92)	Excellent	.48	.43	Considerable
Complex Reflections	.53 (19–.75)	.91** (.71–.97)	Excellent	.66	.60	Strong
MI Adherent	.70 (.36–.86)	.72** (.31–.90)	Good	.62	.48	Considerable
MI Non-Adherent	.01 (−.20 to .41)	.61** (.14–.85)	Good	.28	.25	Poor

Note: ICC = Intra-Class-Correlation; CI = 95% confidence interval; Kappa (TU) = Time-unit kappa, Kappa (E) = Event-based kappa.
[a] Reprinted with permission from Brueck et al. (2009, p. 47).
[b] Categorization based on Cicchetti (1994) for ICCs.
[c] Categorization based on Sachs (1999) for average of Kappa (TU) and Kappa (E).

<.40, poor, .41–.61, considerable, .61–.80, strong, .81–1.00, approaches perfect agreement. For details on kappa classification, see Bakeman and Quera (2011).

Kappa coefficients for the software MITI-d ranged from considerable to strong (time-unit kappas from .48 for Simple Reflections to .83 for Giving Information), except for MI Non-Adherent behavior (time-unit kappa of .28).

3.2. Convergent validity of the software MITI-d

To assess the convergent validity of the MITI-d software version, we obtained ratings of the REM empathy and confrontation subscales for 22 independent video recorded sessions for each rater. We chose this procedure in order to compare our results with a study from Nicolai et al. (2007) in which a similar sample size, but the paper–pencil version of the MITI, were used. Convergent validity was established by comparing MITI-d summary scores with the REM.

The first four columns of Table 4 show results of the convergent validity study from Nicolai et al. (2007). The authors reported the correlations for four summary codes of the MITI (total Reflections, MI Adherence, MI Non-Adherence, and the ratio of Open to Closed Questions) with the REM subscales.

The last four columns of Table 4 show the results of our convergent validity using the software version. In order to make comparisons between the studies easier, we depict the results in the same way as Nicolai et al. did. In both studies, total Reflections were significantly and positively associated with empathy and negatively correlated with confrontation. The reverse pattern was observed for MI Non-Adherence. In both studies, no significant correlations between the MI Adherence and either empathy or confrontation were found. Nonetheless, in both studies, the ratio of Open to Closed Questions showed moderate correlations with empathy and small negative correlations with confrontation scale—yet none of these correlations were significant. Overall, the software version revealed convergent validity results very similar to those reported for the paper–pencil version by Nicolai et al. (2007).

3.3. Behavior slicing

In order to be included in the analysis, video recordings had to last 20 min or more. If an interviewer contributed several recordings, we selected the longest session. Twenty-six independent sessions ($n = 26$) met the criteria, with an average duration of 45.92 min ($SD = 19.83$). We first tested whether a slice, or subsample, containing the first 20 min provided an accurate estimate of the whole session. We computed ICCs between summary scores obtained for the subsample and the same summary scores for the session. The ICCs for all codes were greater than .83; hence, we concluded that using just a subsample of 20 min can result in accurately classifying MITI-specific behaviors.

Table 5
ICCs between summary scores for the whole session (code rates per 5 min) and summary scores for four different 5-min slices and for two different 10-min slices sampled within the first 20 min (code frequencies).

MI codes	5-Min slices				10-Min slices	
	0–5	5–10	10–15	15–20	0–10	10–20
Giving Information	**.76**	**.85**	**.79**	**.65**	**.90**	**.83**
Closed Question	*.55*	**.67**	**.73**	**.63**	**.74**	**.88**
Open Question	*.47*	**.65**	*.55*	**.72**	**.69**	**.79**
Simple Reflection	*.42*	*.53*	**.63**	**.67**	**.69**	**.78**
Complex Reflection	**.77**	**.62**	**.66**	**.71**	**.82**	**.82**
MI Adherent	.15	*.51*	*.54*	*.59*	*.56*	**.65**
MI Non-Adherent	.29	*.58*	**.81**	*.47*	*.58*	**.90**

Note: Values in bold indicate that the slice provides a good estimate of the summary score for the whole session; values in italics indicate that it provides a fair estimate.

We derived two summary scores for the whole session and six summary scores for the slices. Summary scores for the whole session were code rates per 5 and 10 min (R_5 and R_{10}, respectively); for the slices, they were code frequencies for the first, second, third and fourth 5-min segments ($r_{5\text{-}1}$, $r_{5\text{-}2}$, $r_{5\text{-}3}$ and $r_{5\text{-}4}$, respectively) when 5-min slices were explored, and code frequencies for the first and second 10-min segments ($r_{10\text{-}1}$ and $r_{10\text{-}2}$, respectively) when 10-min slices were used.

ICCs between R_5 and $r_{5\text{-}1}$, $r_{5\text{-}2}$, $r_{5\text{-}3,}$ and $r_{5\text{-}4}$ are shown in Table 5. All ICCs are above .40—except for MI adherent and non-adherent sampled within the first 5 min of the session—indicating that 5-min slices provide a fair estimate of the entire session.

ICCs between R_{10} and $r_{10\text{-}1}$ and $r_{10\text{-}2}$ are also shown in Table 5. All ICCs are above the cutoff criterion of .60—except for MI adherent and non-adherent sampled within the first 10 min of the session—therefore providing a good estimate for the entire session.

4. Discussion

The current study presented the use of a software version of a MI coding instrument. Our reliability study indicates that the software version of the MITI-d can be reliably applied by observers who have a reasonable amount of training. The reliability estimates from our study provide initial evidence that the software and paper–pencil version of the MITI-d will yield equivalent results. In addition to this, our results indicate that observers who use the software achieve slightly higher agreements for most codes than using the paper–pencil version. This result can be explained because the software version gives observers better technical means to allocate codes to utterances on the video recordings. The software version allows observers to recode specific events in the interaction stream (Mangold, 2010). More importantly, it enables observers to revise only specific codes (e.g., only the code Closed Questions) without having to go through the entire session. For example, an observer has coded one session but wants to check if he/she correctly coded

Table 4
Correlations between both the paper–pencil MITI-d and the software version with REM scores.

	Paper-and-pencil MITI-d[a]				Software MITI-d			
	Rater 1		Rater 2		Rater 1		Rater 2	
	REM-E	REM-C	REM-E	REM-C	REM-E	REM-C	REM-E	REM-C
MITI total Reflections (rate)	.71**	−.46*	.53**	−.33	.58**	−.48*	.51*	−.18
MITI MI Adherence (rate)	−.07	−.04	.24	−.34	−.15	.06	−.02	.16
MITI MI Non-Adherence (rate)	−.41*	.77**	−.16	.44*	−.81**	.97**	−.70**	.68**
MITI ratio Open/closed Questions	.40	−.22	.15	−.22	.31	−.23	.34	−.17

Note: MITI = Motivational Interviewing Treatment Integrity code; REM = Rating Scales for the Assessment of Empathic Communication; REM-E = Subscale Empathy; REM-C = Subscale Confrontation.
* $p < .05$.
** $p < .01$.
[a] Reprinted with permission from Nicolai et al. (2007, p. 372).

Closed vs. Open Questions. If the session contains only 10% questions, the observer only needs to check a tenth of the entire session. By contrast, the paper–pencil version does not provide such a high-grained resolution of the interaction, and the observer would need to revise the entire session.

Second, we also provided estimates for the point-by-point agreement between both observers. As expected, the point-by-point agreement was lower than the observer agreement for the summary scores of the behavioral codes. Our study showed that the software version can be used reliably for performing subsequent sequential analyses of interviewer behavior (except for the code MI non-adherent behavior).

As the software version preserves the dynamic information of the coded data, future studies could apply it to test specific sequential profiles for interviewers; for example, to examine whether interviewers follow the MI guidelines given by Miller (2000), who recommended that a skilled interviewer should reflect twice for each question he or she asks. This should result in a behavior sequence, such as *question–reflection–reflection*, but should not result in either *reflection–question–reflection* or in *reflection–reflection–question*. Testing whether those behavior sequences actually occur can only be done by means of the software version.

Second, coded video data obtained with the software version can be played back easily to clients in order to verify whether reflections had high empathic accuracy (Moyers & Miller, 2013), for example, by asking clients how much a reflective listening actually was an accurate summary of what the client has said.

Finally, researchers who use other MI instruments (e.g., the MI skill code, Baer et al., 2008; Hodgins, Ching, & McEwen, 2009, Miller, Moyers, Ernst, & Amrhein, 2008) to assess verbal responses of clients within an interaction may easily combine these timed-event data with coded verbal interviewer behavior. By contrast, the paper–pencil version would not allow one to make these comparisons. A combination of different coding instruments for clients and interviewers would allow one to analyze the interactional dynamics of MI (cf., Bakeman & Quera, 2011).

We showed that the software version has convergent validity with the Rating Scales of Empathic Communication. We obtained results similar to those reported in a previous study by Nicolai et al. (2007), who used the paper–pencil MITI-d, and thus we can conclude that the software and the paper–pencil versions provide fairly equivalent behavioral measures.

Finally, we demonstrated how the software allows one to extract behavior slices of the interaction and showed that only a fraction of the interaction is an accurate sample for interviewer behavior across the whole session. Often, researchers have limited resources and cannot code the entire sample (e.g., Brueck et al., 2009; D'Amico et al., 2012). As a consequence, coding only a fraction of the entire session might be a time and money saving endeavor. Our results showed that 10-min slices were accurate samples for MI behaviors, whereas 5-min slices were less accurate. Particularly, we do not recommend using the first 5-min slice from the beginning of the session because both MI Adherent and Non-Adherent codes showed unstable ICCs. We assume that interviewers typically warm up with the client in the first few minutes and therefore do not typically adhere strictly to MI protocol. However, this first preliminary study using the slicing technique of the software instrument showed that up 50% of the coding work can be economized and that behavior estimates were still good to fair. If researchers want to assess MI skills of 185 interviewers (e.g., as in the MI study from Moyers et al., 2005) who provided a 1-h MI session, they could save up to $2158 by coding only 10 instead of 60 min (assuming an average wage of $14 per hour per observer). If these researchers also wanted to extract sequential information, they would need to add another $18,500 for creating transcripts of the sessions (assuming $100 per session; W. Hannöver, personal communication, March 5, 2014). Depending on the research question, the software could have saved between $2000 and $20,000 in this example.

The following limitations need to be considered. First, we obtained our results with a non-clinical student sample. Within our research project, interviewers talked with their clients about changing environmental behavior. Traditionally, MI is a treatment method to reduce substance abuse behavior. However, the intervention is increasingly used for changing other target behaviors, including environmental behaviors (Forsberg, Wickström, & Källmén, 2014; Scalgia, 2014; Tribble, 2008; Wickström, Forsberg, Lindquist, & Källmén, 2011). Both types of intervention

Table A1
Excerpt transcript from a conversation in the MI group. Transcripts have been edited to improve readability.

Speaker	Transcript	MITI code
CP:	Well, in principle, we do it (*caring for the environment*) for conscience' sake... But I think there are a lot of situations where we don't do it right	
I:	You kind of see a lot of potential in yourself to work on these issues	Complex Reflection
CP:	Yeah, definitely	
I:	Why do you do it in some situations (*caring for the environment*)—you have said you do it for your conscience' sake. Is there something beyond that?	Open Question
CP:	Well, you know, it is wrong *not* to be pro-environmental. Mostly for the future, for future generations. For their advantage. They should benefit from our environment. If we waste everything, then there will be nothing left	
I:	It looks like you have a strong feeling of responsibility	Complex Reflection
CP:	Yeah, I don't know if that's so particularly strong. I think that—well I am glad that people who lived before me also thought about it	
I:	You would like to pass this on	Simple Reflection
CP:	Yes	
I:	If somebody else did this for you before, then you could at least pass this on—it is possible	Complex Reflection
CP:	Yes	
I:	Okay, so you have started a little bit and said that you see possibilities in which you could improve	Simple Reflection
I:	Now, I would like to go through some things with you and then we can look together to see if we find something where you say, "oh that is something where I could do something more." Let's start with the first point here: Recycling. You have indicated that you do this sometimes	Giving Information
I:	What do you think about it?	Open Question
CP:	I am not entirely sure. In our household we recycle—and do this properly—and there is also waste-paper—we also recycle that. You know, I don't think it is a big deal to do this	
I:	You do recycle and you do not think that it is asking too much to do this	Simple Reflection
CP:	No, I don't think that it is asking too much	

Note: I = Interviewer; CP = Conversational partner.

Table B1
Decision rules specified in the MITI manual (from Moyers et al., 2003).

MITI codes	Decision rules
Giving Information	Reviewing information contained on assessment instruments does not typically qualify as a reflection, although the reflection code MAY be given if the interviewer skillfully emphasizes or enriches the material the client has given "You indicated on the questionnaire that you currently use your car more than your bike." (Giving Information) Giving information should not be confused with giving advice, warning, confronting, or directing
Open Question	"Tell me more" statements are coded as open questions unless the tone and context clearly indicate a confrontation (MI Non-Adherent)
Closed Question	Occasionally the interviewer will offer a statement that otherwise meets the criteria for a reflection, but is given with an inflection at the end (thereby making it "sound like" a question). These statements are coded as Questions (either open or closed), NOT as reflections
Simple Reflection	When a coder cannot distinguish between a simple and complex reflection, the simple designation should be used. Default category: simple
Complex Reflection	Sometimes the interviewer begins with a reflection, but adds a question to "check" the reliability of the reflection (either open or closed). Both elements should be coded
MI Adherent	The MI Adherent code takes precedence when the utterance *clearly* falls into the MI Adherent category. When in doubt, an alternate code (for example, Open Question or Reflection) should be given
MI Non-Adherent	Restating negative information already known or disclosed by the client can be either a confront or a reflection. Most confrontations can be correctly categorized by careful attention to voice tone and context The MI Non-Adherent code takes precedence when the utterance *clearly* falls into the MI Non-Adherent category. When in doubt, an alternate code (for example, Giving Information) should be given

have in common that interviewers need to motivate clients for behavior change, which is the main focus of MI. Nevertheless, we recommend that other research groups use the software version and replicate our results. Specifically, future research about the adequacy of the slicing method to generalize across clinical samples would be necessary. For these purposes, using the software version highly facilitates extracting behavior slices.

Second, the interviewers in our sample were no therapists, but instead were students who received training in MI. Whereas this may affect the general competency level of MI adherence, this particularity of the sample should not affect the reliability or validity of the observational measures.

5. Conclusion

The present study has presented advantages of a software-supported coding scheme for researchers. Whereas the paper–pencil version only captures event data with no sequence information, the software version automatically records time and sequences of the behavioral code stream. We demonstrated that the software version is a reliable and valid instrument that is equivalent to the paper–pencil version. The advantages of the software version are that it makes revising coded data more easily, it provides estimates of point-by-point reliability, it can be used for sequential analyses, and—depending on the research questions—it can economize data analysis in terms of time and money.

Appendix A

See Table A1.

Appendix B

See Table B1.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chb.2014.10.034.

References

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274. http://dx.doi.org/10.1037/0033-2909.111.2.256.

Baer, J. S., Beadnell, B., Garrett, S. B., Hartzler, B., Wells, E. A., & Peterson, P. L. (2008). Adolescent change language within a brief motivational intervention and substance use outcomes. *Psychology of Addictive Behaviors, 22*, 570. http://dx.doi.org/10.1037/a0013022.

Bakeman, R., Deckner, D. F., & Quera, V. (2005). Analysis of behavioral streams. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 394–420). UK: Oxford.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge, UK: Cambridge University Press.

Bakeman, R., Quera, V., & Gnisci, A. (2009). Observer agreement for timed-event sequential data: A comparison of time-based and event-based algorithms. *Behavior Research Methods, 41*, 137–147. http://dx.doi.org/10.3758/BRM.41.1.137.

Barzilai, S., & Zohar, A. (2006). How does information technology shape thinking? *Thinking Skills and Creativity, 1*(2), 130–145. http://dx.doi.org/10.1016/j.tsc.2006.08.001.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396–403. http://dx.doi.org/10.1111/j.1745-6916.2007.00051.x.

Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior, 23*, 463–477. http://dx.doi.org/10.1016/j.chb.2004.10.020.

Brueck, R. K., Frick, K. M., & Loessl, B. (2006). *Der Motivational Interviewing Treatment Integrity Code, deutsche Version (MITI-d) [The motivational interviewing treatment integrity code, German version (MITI-d)]*. Unpublished manual.

Brueck, R. K., Frick, K., Loessl, B., Kriston, L., Schondelmaier, S., Go, C., et al. (2009). Psychometric properties of the German version of the motivational interviewing treatment integrity code. *Journal of Substance Abuse Treatment, 36*, 44–48. http://dx.doi.org/10.1016/j.jsat.2008.04.004.

Butler, L. H., & Correia, C. J. (2009). Brief alcohol intervention with college student drinkers: Face-to-face versus computerized feedback. *Psychology of Addictive Behaviors, 23*, 163. http://dx.doi.org/10.1037/a0014892.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. http://dx.doi.org/10.1037//1040-3590.6.4.284.

Clayman, M. L., Makoul, G., Harper, M. M., Koby, D. G., & Williams, A. R. (2012). Development of a shared decision making coding system for analysis of patient–healthcare provider encounters. *Patient Education and Counseling, 88*, 367–372.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. http://dx.doi.org/10.1177/001316446002000104.

Daeppen, J. B., Bertholet, N., Gmel, G., & Gaume, J. (2007). Communication during brief intervention, intention to change, and outcome. *Substance Abuse, 28*, 43–51. http://dx.doi.org/10.1300/J465v28n03_05.

D'Amico, E. J., Osilla, K. C., Miles, J. N., Ewing, B., Sullivan, K., Katz, K., et al. (2012). Assessing motivational interviewing integrity for group interventions with adolescents. *Psychology of Addictive Behaviors, 26*, 994. http://dx.doi.org/10.1037/a0027987.

Demmel, R., & Peltenburg, M. (2006). *Motivational interviewing: Kommunikation auf gleicher Augenhöhe [Motivational interviewing: Communication on a level playing field] [DVD]*. Retailer: Neuland.

Feigelson, M. E., & Dwight, S. A. (2000). Can asking questions by computer improve the candidness of responding? A meta-analytic perspective. *Consulting Psychology Journal: Practice and Research, 52*, 248–255. http://dx.doi.org/10.1037//1061-4087.52.4.248.

Finegan, J. E., & Allen, N. J. (1994). Computerized and written questionnaires: Are they equivalent? *Computers in Human Behavior, 10*, 483–496. http://dx.doi.org/10.1016/0747-5632(94)90042-6.

Fleiss, J. L., & Shrout, P. E. (1978). Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika, 43*, 259–262. http://dx.doi.org/10.1007/BF02293867.

Forsberg, L., Wickström, H., & Källmén, H. (2014). Motivational interviewing may facilitate professional interactions with inspectees during environmental inspections and enforcement conversations. *PeerJ, 2*, e508.

Gaume, J., Gmel, G., Faouzi, M., & Daeppen, J.-B. (2008). Counsellor behaviours and patient language during brief motivational interventions: A sequential analysis of speech. *Addiction, 103*, 1793–1800. http://dx.doi.org/10.1111/j.1360-0443.2008.02337.x.

Glynn, L. H., Hallgren, K. A., Houck, J. M., & Moyers, T. B. (2012). CACTI: free, open-source software for the sequential coding of behavioral interactions. *PloS one, 7*(7), e39740. http://dx.doi.org/10.1371/journal.pone.0039740.

Graf, E.-M., Aksu, Y., & Rettinger, S. (2010). Qualitativ-diskursanalytische Erforschung von Coaching-Gesprächen [Qualitative-discourse analytic research on coaching interaction]. *Organisationsberatung, Supervision, Coaching, 17*, 133–149. http://dx.doi.org/10.1007/s11613-010-0188-7.

Gunnar, M. R., Kryzer, E., Phillips, D., & Vandell, D. (2001). *Modified observational ratings of caregiving environment (M-ORCE)*. Unpublished coding instrument and manual.

Hannöver, W., Blaut, C., Kniehase, C., Martin, T., & Hannich, H. J. (2013). Interobserver agreement of the German translation of the Motivational Interviewing Sequential Code for Observing Process Exchanges (MI-SCOPE; D). *Psychology of Addictive Behaviors, 27*, 1196–1200. http://dx.doi.org/10.1037/a0033041.

Hodgins, D. C., Ching, L. E., & McEwen, J. (2009). Strength of commitment language in motivational interviewing and gambling outcomes. *Psychology of Addictive Behaviors, 23*, 122. http://dx.doi.org/10.1037/a0013010.

Houck, J. M., Moyers, T. B., & Tesche, C. D. (2013). Through a glass darkly: Some insights on change talk via magnetoencephalography. *Psychology of Addictive Behaviors, 27*, 489. http://dx.doi.org/10.1037/a0029896.

Jonassen, D. H. (1995). Computers as cognitive tools: Learning with technology, not from technology. *Journal of Computing in Higher Education, 6*(2), 40–73.

Jung, M., Chong, J., & Leifer, L. (2012). Group hedonic balance and pair programming performance: Affective interaction dynamics as indicators of performance. In *Proceedings of the 2012 ACM annual conference on human factors in computing systems* (pp. 829–838). ACM.

La France, B. H., Heisel, A. D., & Beatty, M. J. (2004). Is there empirical evidence for a nonverbal profile of extraversion: A meta-analysis and critique of the literature. *Communication Monographs, 71*, 28–48. http://dx.doi.org/10.1080/03634520410001693148.

La France, B. H., Heisel, A. D., & Beatty, M. J. (2007). A test of the cognitive load hypothesis: Investigating the impact of number of nonverbal cues coded and length of coding session on observer accuracy. *Communication Reports, 20*(1), 11–23. http://dx.doi.org/10.1080/08934210601180739.

Lalouschek, J., & Menz, F. (2002). Empirische Datenerhebung und Authentizität von Gesprächen [Empirical data acquisition and authenticity of conversations]. In G. Brünner, R. Fiehler, & W. Kindt (Eds.), *Angewandte Diskursforschung [Applied discourse research]* (pp. 46–68). Radolfzell: Verlag für Gesprächsforschung.

Madson, M. B., Campbell, T. C., Barrett, D. E., Brondino, M. J., & Melchert, T. P. (2005). Development of the motivational interviewing supervision and training scale. *Psychology of Addictive Behaviors, 19*, 303. http://dx.doi.org/10.1037/0893-164X.19.3.303.

Mangold (2010). *INTERACT quick start manual V2.4. Mangold International GmbH*. <www.mangold-international.com>.

Martin, T., Christopher, P. J., Houck, J. M., & Moyers, T. B. (2011). The structure of client language and drinking outcomes in project match. *Psychology of Addictive Behaviors, 25*, 439. http://dx.doi.org/10.1037/a0023129.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30–46. http://dx.doi.org/10.1037//1082-989X.1.1.30.

Meinecke, A. L., & Lehmann-Willenbrock, N. (in press). Social dynamics at work: Meetings as a gateway. In J. A. Allen, N. Lehmann-Willenbrock, & S. G. Rogelberg (Eds.), *The Cambridge handbook of meeting science*. New York, NY: Cambridge University Press.

Miller, W. R. (2000). *Motivational interviewing skill code (MISC). Coder's manual*. University of New Mexico, Center on Alcoholism, Substance Abuse and Addictions (CASAA). <http://casaa.unm.edu/download/misc1.pdf> Retrieved 17.12.12.

Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). *Manual for the motivational interviewing skill code (MISC): Version 2.1*. <http://casaa.unm.edu/download/misc.pdf> Retrieved 17.12.12.

Miller, W. R., & Moyers, T. B. (2006). Eight stages in learning motivational interviewing. *Journal of Teaching in the Addictions, 5*, 3–17. http://dx.doi.org/10.1300/J188v05n01_02.

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

Motivational Interviewing Network of Trainers (2008). *Motivational interviewing: Training for new trainers. Resources for trainers*. <http://www.motivationalinterview.org/Documents/TNT_Manual_Nov_08.pdf>.

Moyers, T. B., Martin, T., Manuel, J. K., & Miller, W. R. (2003). *The motivational interviewing treatment integrity (MITI) code: Version 2.0*. Unpublished manuscript. Albuquerque, NM: University of New Mexico, Center on Alcoholism, Substance Abuse and Addictions.

Moyers, T. B., & Martin, T. (2006). Therapist influence on client language during motivational interviewing sessions. *Journal of Substance Abuse Treatment, 30*, 245–251. http://dx.doi.org/10.1016/j.jsat.2005.12.003.

Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy, 31*, 177–184. http://dx.doi.org/10.1017/S1352465803002054.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment, 28*, 19–26. http://dx.doi.org/10.1016/j.jsat.2004.11.001.

Moyers, T. B., & Miller, W. R. (2013). Is low therapist empathy toxic? *Psychology of Addictive Behavior, 27*(3). http://dx.doi.org/10.1037/a0030274.

Murphy, J. G., Dennhardt, A. A., Skidmore, J. R., Martens, M. P., & McDevitt-Murphy, M. E. (2010). Computerized versus motivational interviewing alcohol interventions: Impact on discrepancy, motivation, and drinking. *Psychology of Addictive Behaviors, 24*, 628. http://dx.doi.org/10.1037/a0021347.

Nicolai, J., Demmel, R., & Hagen, J. (2007). Rating scales for the assessment of empathic communication in medical interviews (REM): Scale development, reliability, and validity. *Journal of Clinical Psychology in Medical Settings, 14*, 367–375. http://dx.doi.org/10.1007/s10880-007-9081-8.

Peterson, P. L., Baer, J. S., Wells, E. A., Ginzler, J. A., & Garrett, S. B. (2006). Short-term effects of a brief motivational intervention to reduce alcohol and drug risk among homeless adolescents. *Psychology of Addictive Behaviors, 20*, 254. http://dx.doi.org/10.1037/0893-164X.20.3.254.

Project MILES (2011). <http://www.medizin.uni-greifswald.de/medpsych/index.php?id=453>.

Quera, V., Bakeman, R., & Gnisci, A. (2007). Observer agreement for event sequences: Methods and software for sequence alignment and reliability estimates. *Behavior Research Methods, 39*, 39–49. http://dx.doi.org/10.3758/BF03192842.

Rammstedt, B., Holzinger, B., & Rammsayer, T. (2004). Zur Äquivalenz der Papier-Bleistift-und einer computergestützten Version des NEO-Fünf-Faktoren-Inventars (NEO-FFI) [Comparison of paper-and-pencil vs. computerized administration of the NEO-Five-Factor-Inventory (NEO-FFI)]. *Diagnostica, 50*, 88–97. http://dx.doi.org/10.1026/0012-1924.50.2.88.

Rosenfeld, P., Giacalone, R. A., Knouse, S. B., Doherty, L. M., Mitchell Vicino, S., Kantor, J., et al. (1991). Impression management, candor, and microcomputer-based organizational surveys: An individual differences approach. *Computers in Human Behavior, 7*, 23–32.

Sachs, L. (1999). *Angewandte Statistik: Anwendung statistischer Methoden [Applied statistics: Application of statistical methods]* (9th ed.). Berlin: Springer.

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20*(3), 2–9. http://dx.doi.org/10.3102/0013189X020003002.

Scalgia (2014). Motivational Interviewing and environmental education. *Presentation at the international conference on motivational interviewing*, June 16th–18th, Amsterdam.

Serowik, K. L., Ablondi, K., Black, A. C., & Rosen, M. I. (2014). Developing a benefits counseling website for veterans using motivational interviewing techniques. *Computers in Human Behavior, 37*, 26–30. http://dx.doi.org/10.1016/j.chb.2014.03.019.

Sharp, W. G., & Hargrove, D. S. (2004). Emotional expression and modality: An analysis of affective arousal and linguistic output in a computer vs. paper paradigm. *Computers in Human Behavior, 20*, 461–475. http://dx.doi.org/10.1016/j.chb.2003.10.007.

Tapp, J., Ticha, R., Kryzer, E., Gustafson, M., Gunnar, M. R., & Symons, F. J. (2006). Comparing observational software with paper and pencil for time-sampled data: A field test of interval manager (INTMAN). *Behavior Research Methods, 38*, 165–169. http://dx.doi.org/10.3758/BF03192763.

Tribble, S. L. (2008). Promoting environmentally responsible behaviors using motivational interviewing techniques. *Honors Project*. Paper 124. <http://digitalcommons.iwu.edu/psych_honproj/124>.

Vader, A. M., Walters, S. T., Prabhu, G. C., Houck, J. M., & Field, C. A. (2010). The language of motivational interviewing and feedback: Counselor language, client language, and client drinking outcomes. *Psychology of Addictive Behaviors, 24*, 190. http://dx.doi.org/10.1037/a0018749.

Waller, M. J., Sohrab, G., & Ma, B. W. (2013). Beyond 12 angry men: Thin-slicing film to illustrate group dynamics. *Small Group Research, 44*, 1–20. http://dx.doi.org/10.1177/1046496413487409.

Wästlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior, 21*(2), 377–394. http://dx.doi.org/10.1016/j.chb.2004.02.007.

Wickström, H., Forsberg, L., Lindquist, S., & Källmén (2011). *Feasibility of motivational interviewing to promote a positive environmental behavior*. <http://www.effektivmiljotillsyn.se/emt/wp-content/uploads/2011/09/emt-handout-110906-final.pdf>.

Yoder, P. J., & Symons, F. J. (2010). *Observational measurement of behavior*. New York: Springer Publishing Company.