Measuring the quality of motivational interviewing in primary health care encounters: The development and validation of the motivational interviewing assessment scale (MIAS)

Manuel Campiñez Navarro, Luis Ángel Péruola de Torres, Josep M. Bosch Fontcuberta, Nieves Barragán Brun, Juan Carlos Arbonies Ortiz, Jesus Manuel Novo Rodríguez, Julia Bóveda Fontán, Remedios Martín Alvarez, Jose Antonio Prados Castillejo, Gabriela Renée Rivas Doutreleau, Carmen Domingo Peña, Jaime Jesus Castro Moreno and Esperanza Maria Romero Rodríguez

ABSTRACT
Background: Motivational interviewing (MI) is a collaborative, goal-oriented method to help patients change behaviour. Tools that are often used to measure MI are the motivational interviewing skills code (MISC), the ‘motivational interviewing treatment integrity’ (MITI) and the ‘behaviour change counselling index’ (BECCI). The first two instruments have not been designed to be used in primary healthcare (PHC) settings. The BECCI actually is time-consuming. The motivational interviewing assessment scale (MIAS, ‘EVEM’ in Spanish) was developed to measure MI in PHC encounters as an alternative to the previous instruments.

Objectives: To validate MIAS as an instrument to assess the quality of MI in PHC settings.

Methods: (a) Development: Sixteen experts in MI participated in the design, face and consensus validity, using a Delphi-type methodology. (b) Validation: Setting: 27 PHC centres located in Spain. Subjects: four experts in MI tested its psychometric properties with 332 video recordings coming from the Dislip-EM study (consultations provided by 37 practitioners). Measurements: dimensionality, internal consistency, reliability (intra-class correlation coefficient—ICC), sensitivity to change and convergent validity with the BECCI scale.

Results: A 14-item scale was obtained after the validation process. Factor analysis: two factors explained 76.6% of the total variance. Internal consistency, α = 0.99. Reliability: intra-rater ICC = 0.96; inter-rater ICC = 0.97. Sensitivity to change: means before and after training were 23.63 versus 38.57 (P < 0.001). Spearman’s coefficient between the MIAS and the BECCI scale was 0.98 (P < 0.001).

Conclusion: The MIAS is a consistent and reliable instrument to assess the use of MI in PHC settings.

KEY MESSAGES
• The motivational interviewing assessment scale is a valid tool to assess motivational interviewing skills in physicians.
• It has been tested in a wide variety of clinical situations in primary healthcare settings and can also be used as a tool for evaluation of training.

Introduction
Motivational interviewing (MI) is a collaborative, goal-oriented method of communication with particular attention to the language of change. It is designed to strengthen an individual’s motivation for and movement toward a specific goal by eliciting and exploring the person’s arguments for change.

CONTACT Luis Ángel Péruola de Torres luisangel.perula@gmail.com Unidad Docente de Medicina Familiar y Comunitaria de Córdoba, Distrito Sanitario Córdoba y Guadalquivir, Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC)/Hospital Universitario Reina Sofia/Universidad de Córdoba, Avda Menéndez Pidal, s/n. Córdoba, Spain

© 2016 WONCA Europe (the European Society of General Practice/Family Medicine)
Ensuring fidelity to the MI intervention is one of the puzzles that researchers face when assessing the quality of clinical trials. Otherwise, conclusions on the particular results could misrepresent the impact of MI. Several tools have been constructed, developed, and validated to assess the integrity of MI when delivered in clinical encounters. The most commonly used are the ‘motivational interviewing skills code’ (MISC),[2] and the ‘motivational interviewing treatment integrity’ (MITI).[3] These tools have not been designed to provide information on the use of MI in primary healthcare (PHC) settings.

In PHC encounters, time is an issue. In Spain, where the development of the Measuring Motivational Interviewing Skills Scale (MIAS in English; ‘Escala para la Valoración de la Entrevista Motivacional’, EVEM scale, in Spanish) was completed,[4] practitioners are usually not able to spend more than 10 min in consultations with their patients.[5] The MITI advises 20-min samples of encounters when using scale per the author’s methodology.[6] Therefore when trying to evaluate the clinician’s abilities in the use of MI by just counting the therapist’s behaviours (e.g. number of reflections or closed questions), the scores may indicate a low performance in MI using traditional scales, while the real performance may meet MI standards if we pay attention to ‘what is the effect of what has been said in the patient’ instead of ‘how many times it was said’. Additionally, it is very common for patients to introduce other health problems that must be addressed by the professional during the consultation. These interactions may lead to more closed-ended questions than open-ended questions in these interviews, but the quality of the open-ended questions may lead to more change and commitment talk, and thus to change.[7]

In PHC encounters, nonverbal behaviour plays a very important role in the development of the interview. The previously mentioned tools were not designed to capture the elements of the nonverbal behaviour, which we understand that are nuclear to the patient–physician relationship and the development of the conversations about change.[8]

The ‘behaviour change counselling index’ (BECCI),[9] pays special attention to PHC encounters, although it does not evaluate multi-target interventions or nonverbal behaviour, and it assesses specifically behaviour change counselling rather than pure MI. However, we have chosen the BECCI scale in this study to evaluate the convergent validity with the MIAS because of our previous experience in its use and the availability of a Spanish version of the scale.[10,11]

Given that not paying attention to time restraints, multi-target interventions and nonverbal behaviour could lead to underestimating the physicians’ skills in MI when using the previous instruments, the MIAS was constructed.

Methods

Study design

Methods have been published in detail elsewhere.[4] We have detailed in Box 1, the processes and measurements that were tested with each version of the scale. Differences in sample sizes between the original protocol and the validation process were due to feasibility criteria and losses in the development of the Dislip-EM study.[12,13] The Dislip-EM study was a multi-centric randomized control trial conducted in Spain where MI was used in the intervention group to help patients reduce their cardiovascular risk.

The Ethics Committee for Clinical Research of Cordoba approved the study on 30 May 2013 (ref. 2243).

Development phase

Four experts in MI constructed the first version of the scale, coming from an existing tool that measures patient-centeredness, and took its name from the words connect (conectar in Spanish), identify and understand the problems (identificar, comprender), agree (acordar) and help (ayudar): the CICAA questionnaire.[14] A Delphi-based consensus of experts in MI participated in three rounds (n = 16).

Validation phase

The Dislip-EM study provided a wide range of clinical situations, which were recorded and scored with the MIAS. It included 37 physicians from 27 PHC centres in four regions of Spain.

Measurements and statistics

Initial inter-rater pilot tests were made with the earlier versions of the scale (MIAS 1.0 to MIAS 1.3), with small samples (n = 18), to refine the scale.

Intra-rater reliability. Two experts in MI in PHC settings evaluated a sample of 20 interviews coming from the Dislip-EM study with the MIAS 1.3. Interviews for the reliability tests were selected by an independent researcher, who evaluated the quality of MI delivered.
as poor, sufficient or excellent (1/3 of each approximately).

Four non-MI experts who were trained in the use of the scale evaluated a sample of six interviews with the MIAS 2.0. Kappa indexes (simple concordance index (SCI) when Kappa could not be calculated) and intra-class correlation coefficients (ICC) were calculated.[15]

Inter-rater reliability. Four experts in MI (two of which were trained in the use of the scale) evaluated a sample of the same 22 interviews. The global ICC was calculated after these experts had scored 332 interviews coming from the Dislip-EM study as well.

Internal consistency. Cronbach’s alpha was calculated in four different samples ($n_1 = 18$; $n_2 = 22$; $n_3 = 332$; $n_4 = 6$) of interviews coming from the Dislip-EM study, scored by four experts in MI (samples 1, 2 and 3) and four non-experts (sample 4) with different versions of the scale. Both the intra-rater and the inter-rater reliability were also evaluated with the Bland and Altman method.[16]

Convergent validity with the BECCI was tested with two samples. The first one included clinical interviews with 18 physicians and 12 PHC nurses scored by two experts in MI with the MIAS 1.0. The second one included a sample of 20 interviews coming from the Dislip-EM study scored by one expert in MI with the MIAS 2.0. Spearman’s coefficient was calculated.

**Sensitivity to change**

All interviews obtained in the Dislip-EM study were compared before and after the training in MI, both in the control and the intervention group, 236 with...
standardized patients and 96 with real patients. Mean scores, Student’s t and ANOVA tests were used.[17]

Factor analysis. An exploratory analysis was performed with the scores of the 332 encounters from the Dislip-EM study. Kaiser–Mayer–Olkin index and Barlet’s sphericity test were calculated, as well as the eigenvalues of the correlation matrix and the percentage of the total variance explained by each factor.[15]

Training process
In the earlier versions of the scale, two experts in MI were trained in the use of the MIAS. As with the refining process they became contributors, the training process could not be measured in itself. Later, four non MI-experts were trained in the use of the MIAS 2.0 scale and they individually evaluated six interviews receiving feedback through email and one virtual encounter in Google Hangout. All participants trained in the use of the scale were asked about its easiness, cost, balance, clarity, and handling.

Results
The final scale obtained after the validation process included 14 items out of the initial 29: some were eliminated and some reformulated in the refining process. The original scale had a 0–2 Likert scoring, but the latest versions included a 0–4 Likert scoring, which had better reliability results and the experts considered it more useful.

Intra-rater reliability
The evolution of the Kappa indexes (k), when measuring the intra-rater reliability from the earlier versions of the scale to the latest, is shown in Figure 1. SCI were in the 50–100% range. ICC for the observer 1 was 0.98 (95% confidence interval (95% CI): 0.84–0.99), for the observer 2 was 1.00 (95% CI: 0.99–1.00), for the observer 3 was 0.99 (95% CI: 0.97–0.99) and for the observer 4 was 0.96 (95% CI: 0.73–0.99). The global ICC for the four observers was 0.95 (95% CI: 0.90–0.98).

Inter-rater reliability
The Kappa indexes were $k = 0.21–0.40$ in 37.5%, $k = 0.41–0.60$ in 14.3%, $k = 0.61–0.80$ in 28.6% and $k > 0.81$ in 21.4%. The ICC were ICC <0.30 in 9.0% of the measurements, 0.31–0.50 in 22.0%, 0.51–0.70 in 24.0%, 0.71–0.90 in 29% and >0.91 in 16.0%. The global ICC for the 332 interviews was 0.97 (95% CI: 0.96–0.97). The agreement was found to be better in the experts who developed the scale than in the trained ones (Bland and Altman graphics, Figure 2). In addition, the experts tended to score lower than the trained observers did.

![Figure 1. Evolutive comparison of kappa indexes in the intra-rater reliability measurements with an earlier and the latest version of the MIAS (EVEM in Spanish).](image)
Internal consistency
Cronbach’s alpha maintained a high level in the samples scored by the different experts and non-experts and with different versions of the scale ($\alpha_1 = 0.97; \alpha_2 = 0.97; \alpha_3 = 0.97; \alpha_4 = 0.99$).

Convergent validity
The Spearman’s correlation coefficient between the MIAS 1.0 and the BECCI was 0.84 ($P < 0.001$), and it was 0.98 ($P < 0.001$) with the MIAS 2.0 (Figure 3).

Sensitivity to change
There were no differences in the scores of the non-MI and the MI groups before the training in MI to the intervention group (IG). The impact of the training was measured scoring the encounters with standardized patients with the MIAS 1.4, finding that the mean before the training was 23.63, and after the training 38.57 ($t = 4.549; P < 0.001$) in the IG. In real conditions, the scores lowered compared to that of the standardized patients, but increased from visit 1 (22.51) to visit 4 (24.96; $F = 3.039; P = 0.023$), what meant that the scale was able to detect differences also in the assessment of real consultations.

Factor analysis
Two factors that explained 76.6% of the total variance were found. Items 8 (objectives) and 9 (action plan) showed a higher correlation with factor 2, which lead to identifying this factor as ‘directional’ (Table 1). The rest of items showed a better correlation with factor 1, which we called ‘relational’, referring to the

![Figure 2. Bland and Altman’s graphic for inter-rater reliability.](image)

![Figure 3. Convergent validity between the BECCI and the MIAS.](image)

Table 1. Correlation between the items and the 2 two components of the factors analysis.

| Item                      | Correlation with factor 1 | Correlation with factor 2 |
|---------------------------|---------------------------|---------------------------|
| 1. Empathy                | 0.81                      | 0.32                      |
| 2. Positioning            | 0.71                      | 0.52                      |
| 3. Concordance with positioning | 0.72                    | 0.46                      |
| 4. Open questions         | 0.73                      | 0.52                      |
| 5. Reflective listening   | 0.83                      | 0.30                      |
| 6. Summaries              | 0.56                      | 0.15                      |
| 7. Affirmation            | 0.72                      | -0.09                     |
| 8. Objectives             | 0.22                      | 0.93                      |
| 9. Action plan            | 0.20                      | 0.93                      |
| 10. Discord               | 0.76                      | 0.38                      |
| 11. Evocation             | 0.73                      | 0.54                      |
| 12. Collaboration         | 0.79                      | 0.51                      |
| 13. Autonomy              | 0.79                      | 0.50                      |
| 14. Compassion            | 0.78                      | 0.47                      |

186 M. CAMPIÑEZ NAVARRO ET AL.
called ‘spirit of MI’ (evocation, collaboration, autonomy support and compassion). The global scores in the item-total correlation showed lower scorings in affirmation and summaries, which could explain the lower correlation with factor 1, and higher scoring for ‘spirit of MI’ items and empathy.

Training process

Participants in the training process dedicated five to seven hours to learn how to score with the MIAS. They needed to score six interviews to get high inter-rater reliability (ICC > 0.80). They evaluated the process as being mostly easy (6/10 score), cheap (5.25/10), and rated the scale as well balanced (8.25/10), clear (8.5/10) and easy to handle (8.25/10). An average 30 min period was needed to score an interview.

Discussion

Development of the scale

The process of constructing and developing of the scale has been a continuum of adapting the scale to the changes in the theoretical model of MI, from the CICAA scale (29 items) to the MIAS 2.0 (14 items). The experts in the consensus group mostly had been taught with the first edition of the MI book.[18] As the refining of the scale and the reliability tests were performed, there was an effort to adapt the content of the scale and its manual to the second [19] and third [1] editions of the MI book.[19,1]

This process was also an exercise of simplifying the scale and being more concise with the definitions of the items and the way they were evaluated. This is reflected in the fact that both inter-rater and intra-rater reliability were improved, as well as internal consistency, with the latest version of the scale, even though it was scored by non-MI experts (Figure 1).

Main results of the validation process

The global ICCs for the intra-rater and the inter-rater tests were both above 0.91.[15] Some Kappa indexes with the Likert 0–2 MIAS 1.3 were low (k < 0.40), but there was an improvement in the ICCs for those items with the Likert 0–4 MIAS 1.4. The scale showed an excellent internal consistency in all its versions (α > 0.90). These results do not differ from the ones obtained in the validation processes of the MISC,[2] the MITI or the BECCI.[3,9]

The convergent validity of a previously validated scale that measured the same theoretical construct (BECCI) was almost linear.

The scale has been able to detect changes in the group of physicians who were taught MI before and after the training as well as in their performance with simulated and real patients, so it demonstrates to be useful in evaluating the training process as well as the real consultations.

Distinctive features of the MIAS

The better correlation of the ‘strategies items’ (reflective listening, open questions, affirmation and summaries) with the relational factor in the factor analysis supports our intention to measure the strategies in a qualitative way, rather than just count behaviours of the therapist. Given the example of empathy, we consider nonverbal empathy as a previous condition to a good empathy performance. We suggest scoring low empathy in those interviews where the professional does not maintain eye contact and is more aware of the documents, computer, etc. than the patient’s facial expression and the bidirectionality of the information. We score high empathy when nonverbal empathy is present and the professional does not miss important opportunities to empathize with the patient with a specific sentence. We score medium empathy when there is clearly nonverbal empathy, but the professional misses some chances of empathizing with a sentence trying not to interrupt the conversation. This is an example of how we get together the concepts of time, multi-targeting and nonverbal behaviour in the assessment of MI quality.

The training process showed that it is affordable to learn how to use the scale. Compared to other scales, there is less time needed to learn how to use the MIAS, as well as fewer interviews.[20] Coders have rated high its clarity and ease of use, as well as the process of coding to be ‘cheap’. The scale has been tested mostly in a population of physicians, but also in a small sample of nurses in PHC settings. A wide variety of clinical situations emerged from the interviews scored, not only those referred to ‘target behaviours’. The scale has been conceived in a way that it can detect good performing of MI in a usual PHC consultation, and not only in specific conversations about change.

The scale has been designed to evaluate video recordings of consultations, as it takes nonverbal aspects of communication into account, which becomes a limitation in its use with audiotapes, which are easier to obtain.
**Implications for clinical practice**

The MIAS is useful as an evaluation tool in MI skills to select participants in a research study and as a tool for evaluation of training in PHC settings.

**Conclusion**

The MIAS shows a good reproducibility. It demonstrates its homogeneity and high internal consistency. The factor analysis shows two components—relational and directional—which are correlated with the theoretical model proposed for the MI method. The minimum number of clinical encounters that need to be evaluated by observers to get a good inter-rater reliability is of six interviews. The MIAS shows an excellent convergent validity with the BECCI scale. It also shows sensitivity to change, as it detects changes in the training process and in clinical abilities in real practice when delivering MI in PHC settings.

**Acknowledgements**

Collaborative Group Dislip-EM.

**Declaration of interest**

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

**Funding**

This study was supported by the Spanish Society of Family Physicians (semFYC) and the Health Council of the Andalusian Government (PI-0421-2010).

**References**

[1] Miller WR, Rollnick S. Motivational Interviewing: Helping people change. New York, London: Guilford Press; 2013.

[2] Moyers TB, Martin T, Catley D, et al. Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. Behav Cogn Psychother. 2003;31:177–184.

[3] Moyers TB, Martin T, Manuel JK, et al. Assessing competence in the use of motivational interviewing. J Subst Abuse Treat. 2005;28:19–26.

[4] Pérola LA, Campiñez M, Bosch JM, et al. Is the scale for measuring motivational interviewing skills a valid and reliable instrument for measuring the primary care professionals motivational skills? EVEM study protocol. BMC Fam Pract. 2012;13:112.

[5] Gervas J, Palomo L, Pastor-Sánchez R, et al. Pressing problems in primary care. Aten Primaria 2001;28:472–477.

[6] Moyers TB, Martin T, Manuel JK, et al. Revised global scales: Motivational interviewing treatment integrity 3.1(MITI 3.1.1) [Internet]. 2010 [cited 13 January 2016]. Available from: http://casaa.unm.edu/download/MITI3_1.pdf

[7] Amrhein PC, Miller WR, Yahne CE, et al. Client commitment language during motivational interviewing predicts drug use outcomes. J Consult Clin Psychol. 2003;71:862–878.

[8] Mast MS. On the importance of nonverbal communication in the physician–patient interaction. Patient Educ Couns. 2007;67:315–318.

[9] Lane C, Huws-Thomas M, Hood K, et al. Measuring adaptations of motivational interviewing: The development and validation of the behaviour change counselling index (BECCI). Patient Educ Couns. 2005;56:166–173.

[10] Campiñez M. Formulario del Índice del Consejo sobre el Cambio de Conducta [Internet]. MINT; 2008 [cited 13 January 2016]. Available from: http://www.motivationalinterviewing.org/sites/default/files/BECCI%20-%20Spanish%20Form.pdf

[11] Campiñez M.Manual del Índice del Consejo sobre el Cambio de Conducta. MINT; 2008 [cited 13 January 2016]. Available from: http://www.motivationalinterviewing.org/content/becci-spanish-form

[12] Pérola LA, Bosch JM, Bóveda J, et al. Effectiveness of motivational interviewing in improving lipid level in patients with dyslipidemia assisted by general practitioners: Dislip-EM study protocol. BMC Fam Pract. 2011;12:125.

[13] Bóveda-Fontán J, Barragán-Brun N, Campiñez-Navarro M, et al. Effectiveness of motivational interviewing in patients with dyslipidemia: A randomized cluster trial. BMC Fam Pract. 2015;16:151.

[14] Gavilán Moral E, Ruiz Moral R, Pérola de Torres LA, et al. Evaluation of the patient-centered clinical relationship: Analysis of psychometric properties using the CICAA scale. Aten Primaria 2010;42:162–168.

[15] Carretero-Dios H, Pérez C. Standards for the development and review of instrumental studies. Int J Clin Health Psychol. 2005;5:521–551.

[16] Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurements. Lancet 1986;8476:307–310.

[17] Fritz J. Sensitivity to change. Phys Ther. 1999;79:420–422.

[18] Miller WR, Rollnick S. Motivational interviewing: Preparing people to change addictive. New York: Guilford Press; 1991.

[19] Miller WR, Rollnick S. Motivational Interviewing: Preparing people for change. New York: Guilford Press; 2002.

[20] Forsberg L, Kallmen H, Hermansson U, et al. Coding counsellor behaviour in motivational interviewing sessions: Inter-rater reliability for the Swedish motivational interviewing treatment integrity code (MITI). Cogn Behav Ther. 2007;36:162–169.