

Advancing Motivational Interviewing Training with Artificial Intelligence: ReadMI

Paul J Hershberger¹
 Yong Pei²
 Dean A Bricker³
 Timothy N Crawford^{1,4}
 Ashutosh Shivakumar²
 Miteshkumar Vasoya²
 Raveendra Medaramitta²
 Maria Rehtin⁵
 Aishwarya Bositty²
 Josephine F Wilson⁴

¹Department of Family Medicine, Wright State University Boonshoft School of Medicine, Dayton, OH, USA;

²Department of Computer Science and Engineering, College of Engineering and Computer Science, Wright State University, Dayton, OH, USA;

³Department of Internal Medicine, Wright State University Boonshoft School of Medicine, Dayton, OH, USA;

⁴Department of Population and Public Health Sciences, Wright State University Boonshoft School of Medicine, Dayton, OH, USA; ⁵Boonshoft School of Medicine, Wright State University, Dayton, OH, USA

Background: Motivational interviewing (MI) is an evidence-based, brief interventional approach that has been demonstrated to be highly effective in triggering change in high-risk lifestyle behaviors. MI tends to be underutilized in clinical settings, in part because of limited and ineffective training. To implement MI more widely, there is a critical need to improve the MI training process in a manner that can provide prompt and efficient feedback. Our team has developed and tested a training tool, Real-time Assessment of Dialogue in Motivational Interviewing (ReadMI), that uses natural language processing (NLP) to provide immediate MI metrics and thereby address the need for more effective MI training.

Methods: Metrics produced by the ReadMI tool from transcripts of 48 interviews conducted by medical residents with a simulated patient were examined to identify relationships between physician-speaking time and other MI metrics, including the number of open- and closed-ended questions. In addition, interrater reliability statistics were conducted to determine the accuracy of the ReadMI's analysis of physician responses.

Results: The more time the physician spent talking, the less likely the physician was engaging in MI-consistent interview behaviors ($r = -0.403$, $p = 0.007$), including open-ended questions, reflective statements, or use of a change ruler.

Conclusion: ReadMI produces specific metrics that a trainer can share with a student, resident, or clinician for immediate feedback. Given the time constraints on targeted skill development in health professions training, ReadMI decreases the need to rely on subjective feedback and/or more time-consuming video review to illustrate important teaching points.

Keywords: patient engagement, medical education, software development

Plain Language Summary

Motivational interviewing (MI) is an effective approach for addressing the patient's behavior as a component of chronic disease management, but requires interview skills that are different than simply educating and advising patients. MI training can be time and labor-intensive so that it may get limited attention in the medical education curriculum. We are developing and testing a training tool, Real-time Assessment of Dialogue in Motivational Interviewing (ReadMI), that makes use of artificial intelligence to provide immediate and quantitative feedback in the MI training process on important metrics in the MI approach.

Introduction

Motivational interviewing (MI) is an evidence-based, brief interventional approach that has been demonstrated to be highly effective in triggering change in high-risk lifestyle behaviors. Even for patients with low readiness to change, MI serves as a critical prelude to contemplating change.¹⁻⁶ MI is a patient-focused conversation

Correspondence: Paul J Hershberger
 Department of Family Medicine, Wright State University Boonshoft School of Medicine, Dayton, OH, USA
 Tel +1 937-245-7223
 Fax +1 937-245-7931
 Email paul.hershberger@wright.edu

between a clinician and patient that reinforces the patient's motivation to make positive changes in any targeted health behavior, including risk-factor modifications for adult and pediatric populations with chronic illnesses.^{7–10} Initially used to treat addictions, MI's efficacy has been demonstrated in numerous randomized trials across a range of conditions and settings.¹¹

As a patient-centered approach, MI emphasizes elicitation of the patient's goals, concerns, and perspective on potential behavior change through exploration of the patient's normal and natural ambivalences. However, as a result of their professional training, the natural inclination of many healthcare providers is to take a directive role to educate and instruct the patient with steps to improve health.^{12,13} Therefore, during MI training, clinicians must learn to change this natural inclination to lead the discussion and allow the conversation to be patient-centered. This means the clinician must talk less, listen and reflect what the patient expresses, and ask open-ended questions – critical skills in the MI approach. Box 1 summarizes the basic elements of MI. Real-time feedback is advantageous in this learning process,^{14,15} but many clinicians receive little feedback on interviewing after the early part of their training.¹⁶

Box 1 Motivational Interviewing Elements

Motivational interviewing (MI) is a collaborative conversation style for strengthening a person's own motivation and commitment to change. It is characterized by a spirit of collaboration, acceptance, compassion/empathy, and evocation.

Core skills of MI include:

- Open-ended questions. Ask questions that cannot be answered with "yes" or "no." Begin questions with: "What ... " "How ... " "When ... " "Where ... " "Who ... " "Tell me more about ... "
- Affirmations. Statements that accentuate positive patient attributes or behavior. ("You did even more than you had hoped to accomplish this week!")
- Reflective Listening. Statements that briefly summarize what the patient has said and/or make a guess about the meaning of what the patient is expressing. Reflective statements can be simple (no additional content or meaning) or complex (additional/different meaning added; containing a guess). A reflective statement is what would come after "Do you mean that ... " without the "Do you mean that" preface. Silence after a reflection invites the patient to say more. ("You're very concerned about the possibility of developing diabetes.")
- Summarizing. Combination of several reflections with intent to draw together the patient's concerns, motivations, intentions, and/or plans.
"Change Rulers" are 0–10 scales are used to assess importance, readiness to change, confidence, etc.

Because of limited and ineffective training, MI is under-utilized,^{16,17} and the "educate and advise" approach to patient behavior is reflexively employed. To implement MI widely, there is a critical need to improve the MI training process in a manner that can provide prompt and efficient feedback to learners when their available time is limited. Our team has developed and tested a training tool, Real-time Assessment of Dialogue in Motivational Interviewing (ReadMI), that uses natural language processing (NLP) to provide MI metrics and thereby address the need for more effective MI training.¹⁸

ReadMI is a low cost, ultra-portable solution to enable automatic and instantaneous MI training assessment and analysis with the use of android tablets. It makes use of the latest advances in deep-learning-based speech recognition, NLP, human-artificial intelligence (AI) teaming, and mobile-cloud computing technologies. ReadMI's system architecture includes a) speech recognition; b) speech separation, ie, detecting speech source in the presence of multiple talkers; and c) MI skill assessment. In real-time, ReadMI produces a spectrum of metrics for MI skills evaluation, including the number of open- and closed-ended questions asked, provider versus patient conversation time, number of reflective statements, and use of a change ruler (0–10 scale to rate importance, readiness, or confidence), which are all integral parts of MI, eliminating the need for time-consuming reviews of recorded training sessions.^{18,19}

ReadMI transcriptions are obtained by utilizing the Google Cloud Speech automatic speech recognition (ASR) via the Google Cloud Speech application programming interface (API). We chose Google Cloud Speech ASR service as it has the best accuracy of all competing ASR services with a Word Error Rate (WER) of 4.5%. Next, in the role-play dialogue analysis, ReadMI carries out the automated behavioral coding by a rule-based NLP classifier that identifies the typical lexical cues used by MI experts and labels the doctor utterances into 5 codes/categories: Open-ended question, Closed-ended question, Reflective statement, Use of Change Ruler, and None (physician just talking). Finally, in the assessment dissemination, the numerical counts of the mentioned codes and the clinician speaking time in percentage constitute the real-time feedback by ReadMI, in addition to the full list of utterances shown for each category. Additionally, ReadMI allows the MI trainee or the expert trainer to load and review previous role-plays and their corresponding analysis results. Saving a session and its metrics also

Box 2 Motivational Interviewing Example

A physician is seeing a patient who is requesting an increase in opioid medication for chronic pain.

Physician: "You mentioned that your previous doctor had increased your dose a few times when you said the medicine didn't seem to be working well enough." (Reflective statement)

Patient: "Yes, it seemed like it would work for a while, but then not be as effective."

Physician: "What do you know about a potential side effect of this medication known as hyperalgesia?" (Open question)

Patient: "That's a big word, Doc. I've never heard of it."

Physician: "Would it be OK with you if I explain a bit about it?" (Closed question; getting permission to educate)

Patient: "Sure."

Physician: "When a patient is on this type of medication for a long period of time, it can begin to have the opposite effect, meaning that you experience more rather than less pain." (Statement, not coded by ReadMI)

Patient: "Do you think that's happening to me?"

Physician: "How does this seem to fit with what has been happening for you?" (Open question)

Patient: "I guess it could be. I want less rather than more pain!"

Physician: "Given the concern that this medication may be no longer helping, I am curious how ready you would be to consider a slow decrease in this medication and try other options. Where would you rate yourself on a scale of 0–10, with 0 being not ready at all to make such a change and 10 being ready to start now?" (Open question; scale – change ruler)

helps create new data points for data analysis and facilitate further refinement of ReadMI and its NLP-based behavioral coding classifiers.

Box 2 contains an interview excerpt showing the classification of the physician utterances. ReadMI delivers immediate feedback to the MI facilitator and trainee. Targeted feedback and repeated practice are essential for progress toward optimal MI utilization by providers, and MI has been associated with increased patient activation and engagement in care.^{20–22}

Standard MI training and assessment methods can be time-consuming and labor intensive, and therefore MI often is given limited attention in health professions education. Meaningful feedback and evaluation for trainees on their MI skills often does not occur if interviews must be recorded, transcribed, and analyzed by an expert trainer. Thus, it is difficult for trainees to establish timely awareness of deficiencies in their use of the MI approach, potentially resulting in loss of interest in MI practice.²³ Immediate and targeted feedback is necessary for overriding their directive tendencies and developing

facility in use of the MI approach with patients. By producing and analyzing an interview transcript in real time and promptly providing metrics on core MI skills, ReadMI makes it possible to incorporate meaningful MI training for medical students, and other health professions students and clinicians, even when time is limited.

To date, our team has been testing ReadMI in MI training sessions with residents and medical students. The current version of ReadMI automatically produces complete transcripts of the MI dialogue with over 92% accuracy and yields metrics on provider versus patient conversation time (over 95% accuracy) and the number of open- and closed-ended questions (over 92% accuracy).¹⁸ This paper reports the results of preliminary analyses of ReadMI metrics using transcripts of interviews conducted by residents in Family Medicine and Internal Medicine while receiving MI training.

Methods

Data

Transcripts from 48 interviews conducted by medical residents with a simulated patient were obtained and analyzed with the ReadMI tool. The context for these interviews was a motivational interviewing training module for Internal Medicine and Family Medicine residents. These training modules typically include approximately 70% Internal Medicine residents and 30% Family Medicine residents, and equal numbers of males and females. The training module is further described in the Procedure section below. The study was given an exemption by the institutional review board (IRB) of Wright State University, Dayton, OH as the project involves research on the effectiveness of an instructional technique in a normal educational practice.

Procedure

During a four-week residency rotation in ambulatory care, approximately 4–6 residents in Family Medicine and Internal Medicine participate in four 2.5-hour sessions of MI training. The training includes substantial practice in role plays with each other, and with a simulated patient in the fourth session. The simulated patient presents one of several prepared scenarios to each resident, eg, a patient requesting more opioid pain medication for chronic pain, or a clergy member who was using marijuana to cope with the stress of her job. During their role play with the simulated patient, each resident's interview was recorded and analyzed by the ReadMI app.

MI metrics produced by the ReadMI app were examined to identify relationships between physician-speaking time and other MI metrics, such as the number of open- and closed-ended questions, number of reflective statements, and use of the change ruler. In addition, five MI training facilitators read the transcripts created by the ReadMI app and rated physician responses as closed-ended, open-ended, reflective, scale (ie, change-ruler), or none (when the physician's response was neither closed-ended, open-ended, reflective, nor used a change-ruler). Interrater reliability statistics were conducted to determine the accuracy of the ReadMI app's analysis of physician responses.

Statistical Analysis

Pearson r correlation analyses were conducted to examine relationships between physician-speaking time and the number of open- and closed-ended questions, number of reflective statements, and use of the change ruler. To examine interrater agreement among the five raters and the ReadMI app, a Fleiss' Kappa statistic was computed along with the percent agreement. Fleiss' Kappa is used as index of agreement between more than two raters and when the ratings are categorical. An intraclass correlation coefficient (ICC) and 95% confidence intervals (CI) were computed to examine agreement among the five raters and the ReadMI ratings for the frequency of categories selected. All data were analyzed using R (R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistics software (Version 25; IBM Corp., Armonk, NY, USA) and p values <0.05 were regarded as statistically significant.

Results

Correlation analyses between physician-speaking time and the number of open- and closed-ended questions, number of reflective statements, and use of the change ruler revealed two significant correlations. A significant, negative correlation was found between physician-speaking time and the number of open-ended questions asked, $r = -0.361$, $p = 0.016$. That is, the more time a physician spends talking, the fewer open-ended questions they ask. Similarly, the more time a physician spends talking, the less likely the physician is to engage in MI-consistent interview behaviors (reflective statements, open-ended questions, change ruler), $r = -0.403$, $p = 0.007$.

Table 1 presents the overall Kappa and the Kappa for the type-of-question rating given for each line ($n = 313$) of dialog for each transcript by the five human raters and the ReadMI app. Overall, the Kappa among the five raters and

Table 1 Kappa Statistics Overall and by Type of Question/Statement

	Kappa Statistic (% Agreement)	p-value
Overall (with 5 raters and ReadMI app)	0.502 (33.3)	<0.0001
Type of Question/Statement		
With 6 Raters (5 human raters and ReadMI)		
Open-ended question	0.596 (36.2)	<0.0001
Closed-ended question	0.428 (27.0)	<0.0001
Reflective statement	0.351 (20.3)	<0.0001
Use of Change Ruler	0.793 (40.2)	<0.0001
None (Physician just talking)	0.472 (33.1)	<0.0001

the ReadMI app was 0.502 (percent agreement = 33.3%) ($p < 0.0001$), suggesting moderate agreement. The largest Kappa was for use of the change ruler (Kappa = 0.793; percent agreement = 40.2%; $p < 0.0001$), suggesting substantial agreement among the five human raters and the ReadMI app, followed by identification of open questions (Kappa = 0.596; percent agreement = 36.2%; $p < 0.0001$), suggesting moderate agreement. Least agreement was found for reflective statements. Table 2 presents the Kappa's between each rater. Agreement between each rater ranged from 0.313 (percent agreement = 51.0%), suggesting fair agreement, to 0.658 (percent agreement = 76.8%), suggesting substantial agreement. For each type of question/statement, the frequency of ratings was summed for each rater (Table 3). An ICC was computed to assess agreement among the frequencies. The ICC among the five human raters and the ReadMI app was 0.828 (95% CI = 0.569–0.977). Thus, the ReadMI app analyzed and rated the type of question/statement as well as the human raters.

The qualitative feedback received from residents and medical students using ReadMI has been positive and highlights how quantitative data can complement subjective feedback from training facilitators. For example, seeing that one was actually speaking 68% of the time helps validate a facilitator's observation that the doctor seemed to be doing a large proportion of the talking. Similarly, seeing the actual count of open and closed questions (with examples from the transcript) is a way to illustrate how simply changing the preface of a question (eg, from "Do you have any ideas ..." to "What ideas do you have ...")

Table 2 Kappa Statistics (% Agreement) Between Two Raters

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	ReadMI App
Rater 1	1.00					
Rater 2	0.539 (66.8)	1.00				
Rater 3	0.313 (51.0)	0.658 (76.8)	1.00			
Rater 4	0.462 (62.2)	0.498 (64.6)	0.466 (63.4)	1.00		
Rater 5	0.645 (74.7)	0.642 (74.2)	0.442 (60.7)	0.477 (62.9)	1.00	
ReadMI App	0.417 (57.7)	0.613 (72.4)	0.525 (67.4)	0.401 (57.4)	0.482 (62.4)	1.00

Table 3 Frequency and Percent of Question/Statement Rating

	None	Open-Ended	Closed-Ended	Reflective	Scale
	n (%)	n (%)	n (%)	n (%)	n (%)
Rater 1	70 (23.0)	127 (41.2)	79 (25.9)	15 (4.9)	14 (4.6)
Rater 2	128 (41.0)	94 (30.1)	47 (15.1)	34 (10.9)	9 (2.9)
Rater 3	165 (53.7)	66 (21.5)	34 (11.1)	34 (11.1)	8 (2.6)
Rater 4	108 (34.6)	117 (37.5)	63 (20.2)	13 (4.2)	11 (3.5)
Rater 5	92 (29.6)	106 (34.1)	84 (27.0)	18 (5.8)	11 (3.5)
ReadMI App	124 (39.6)	91 (29.1)	49 (15.7)	40 (12.8)	9 (2.9)

can readily increase the proportion open-ended questions. Given that medical students, residents, and practicing physicians are accustomed to being evaluated with numerical scores, and that much clinical data is reported numerically, using AI to aid in the training of effective patient engagement skills represents an important use of technology. However, since AI cannot capture the overall spirit of MI nor contextual factors for a patient, ReadMI's role in MI training is adjunctive.²⁴

Discussion

In this study, the ReadMI app had similar accuracy as human raters in identifying the types of questions and statements produced by the MI trainee. As expected, physicians who spend more time talking, compared to the patient, are less likely to engage in MI-consistent interview behaviors. These preliminary results demonstrate the significant benefit of making ReadMI results available in real time to the trainer and trainees. In role-play training sessions with a simulated patient, ReadMI produces specific metrics that a trainer can share with the learner for immediate feedback. This can be particularly effective for situations in which the trainee is doing most of the talking and primarily asking closed-ended questions.

Given the time constraints on targeted skill development faced in health professions education, ReadMI decreases the

need to rely on subjective feedback and/or more time-consuming video review to illustrate important teaching points. We are currently testing ReadMI in a randomized controlled trial, comparing medical students and residents who receive MI training with ReadMI technology to those who receive the usual and customary MI training. ReadMI has the potential to improve health professions education and healthcare quality by making providers better equipped as decision support agents in their efforts to help patients change their health-related behavior.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This project was supported by grant number R21HS026548 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the

authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Disclosure

None of the authors have any conflicts of interest to report.

References

1. Miller WR, Rollnick S. *Motivational Interviewing: Helping People Change*. New York: Guilford; 2013.
2. Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. *Addiction*. 2009;104(5):705–715. doi:10.1111/j.1360-0443.2009.02527.x
3. Rubak S, Sandbæk A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. *Br J Gen Pract*. 2005;55(513):305–312.
4. Lundahl B, Burke BL. The effectiveness and applicability of motivational interviewing: a practice-friendly review of four meta-analyses. *J Clin Psychol*. 2009;65:1232–1245. doi:10.1002/jclp.20638
5. Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *J Consult Clin Psychol*. 2003;71(5):843–861. doi:10.1037/0022-006X.71.5.843
6. Douaihy A, Kelly TM, Gold MA, editors. *Motivational Interviewing: A Guide for Medical Trainees*. New York (NY): Oxford University Press; 2016.
7. Sanders KA, Whited A, Martino S. Motivational interviewing for patients with chronic kidney disease. *Semin Dial*. 2013;26(2):175–179. doi:10.1111/sdi.12052
8. Christie D, Channon S. The potential for motivational interviewing to improve outcomes in the management of diabetes and obesity in paediatric and adult populations: a clinical review. *Diabetes Obes Metab*. 2014;16(5):381–387. doi:10.1111/dom.12195
9. Coleman MT, Newton KS. Supporting self-management in patients with chronic illness. *Am Fam Physician*. 2005;72(8):1503–1510.
10. Anderson W. Collaborative care and motivational interviewing: improving depression outcomes through patient empowerment interventions. *Am J Manag Care*. 2007;13:S103–S106.
11. Frost H, Campbell P, Maxwell M, et al. Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: a systematic review of reviews. *PLoS One*. 2018;13:e0204890. doi:10.1371/journal.pone.0204890
12. Noordman J, van der Weijden T, van Dulmen S. Communication-related behavior change techniques used in face-to-face lifestyle interventions in primary care: a systematic review of the literature. *Patient Educ Couns*. 2012;89(2):227–244. doi:10.1016/j.pec.2012.07.006
13. Werner JJ, Lawson PJ, Panaite V, Step MM, Flocke SA. Comparing primary care physicians' smoking cessation counseling techniques to motivational interviewing. *J Addict Med*. 2013;7(2):139–142. doi:10.1097/ADM.0b013e3182879cc5
14. Pollak KI, Nagy P, Bigger J, et al. Effect of teaching motivational interviewing via communication coaching on clinician and patient satisfaction in primary care and pediatric obesity-focused offices. *Patient Educ Couns*. 2016;99(2):300–303. doi:10.1016/j.pec.2015.08.013
15. Barwick MA, Bennett LM, Johnson SN, McGowan J, Moore JE. Training health and mental health professionals in motivational interviewing: a systematic review. *Child Youth Serv Rev*. 2012;34(9):1786–1795. doi:10.1016/j.childyouth.2012.05.012
16. Schwalbe CS, Oh HY, Zweben A. Sustaining motivational interviewing: a meta-analysis of training studies. *Addiction*. 2014;109:1287–1294. doi:10.1111/add.12558
17. Hall K, Staiger PK, Simpson A, Best D, Lubman DI. After 30 years of dissemination, have we achieved sustained practice change in motivational interviewing? *Addiction*. 2015;111:1144–1150. doi:10.1111/add.13014
18. Vasoya MM, Shivakumar A, Pappu S, et al. ReadMI: an innovative app to support training in motivational interviewing. *J Grad Med Ed*. 2019;11(3):344–346. doi:10.4300/JGME-D-18-00839.1
19. Martino S, Paris M, Anez L, et al. The effectiveness and cost of clinical supervision for motivational interviewing: a randomized controlled trial. *J Subst Abuse Treat*. 2016;68:11–23. doi:10.1016/j.jsat.2016.04.005
20. Anderson RM, Funnell MM, Butler PM, Arnold MS, Fitzgerald JT, Feste CC. Patient empowerment. Results of a randomized controlled trial. *Diabetes Care*. 1995;18(7):943–949. doi:10.2337/diacare.18.7.943
21. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. *Health Aff*. 2013;32:207–214. doi:10.1377/hlthaff.2012.1061
22. Hibbard JH, Greene J, Tusler M. Improving the outcomes of disease management by tailoring care to the patient's level of activation. *Am J Manag Care*. 2009;15(6):353–360.
23. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Srikumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. *J Subst Abuse Treat*. 2016;65:43–50. doi:10.1016/j.jsat.2016.01.006
24. Webster CS. Artificial intelligence and the adoption of new technology in medical education. *Med Educ*. 2021;55(1):6–7. doi:10.1111/medu.14409

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress

including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.