RESEARCH BRIEF

Comparing Human and Artificial Intelligence Coding of Motivational Interviewing Skills Among Child Welfare Professionals

Center *for* Practice TRANSFORMATION
UNIVERSITY OF MINNESOTA

Michael Van Wert, MPH, MSW, LICSW,
Emily Gus, MPH

Background

Although MI has the potential to be useful in a wide range of practice settings and applications (Miller, 2023), including in child welfare (Forrester et al., 2018; Chaffin et al., 2009; Forrester et al., 2008; Carroll et al., 2001), the question of how to most effectively and efficiently train and evaluate provider MI skill remains. There are many human-coded MI fidelity measures, including the Motivational Interviewing Skills Code (MISC), the Motivational Interviewing Treatment Integrity Scale (MITI), and the Behavior Change Counseling Index (BECCI) (Gill, Oster, & Lawn, 2020; Madson & Campbell, 2006). The MITI, widely regarded as a "gold standard" for assessing MI proficiency, has been shown to correlate significantly with the BECCI, a briefer and less resource-intensive alternative (Darnell et al., 2019). Moreover, in order to further reduce the cost in time and effort to human coders, artificial intelligence (AI) platforms that evaluate MI provider performance have been paired with human trainer feedback (Hershberger et al., 2024) and within asynchronous trainings (Cheffer, Barnett, & Bullara, 2025) to improve MI skill.

The present brief sought to answer the following questions:

1. To what extent do human-coded BECCI scores correlate with AI-coded Lysnn MI Proficiency scores?
2. To what extent are changes in human-coded BECCI scores from pre- to post-training associated with changes in Lyssn MI Proficiency scores?
3. What Lyssn MI Proficiency cutoff scores match meaningful cutoffs in BECCI scores used to determine various skill levels (e.g., 2 or above for MI certification for Minnesota child welfare professionals)?

Methods

The present data was generated as part of a 28-month pilot to increase child welfare professionals' (CWP) MI skill. The project was conducted by the Center for Practice Transformation (CPT) and the Minnesota Child Welfare Training Academy (MCWTA), and sponsored by the MN Department of Human Services and the Family First Prevention Services Act. During the pilot, some counties chose to attend an intensive training led by CPT and MCWTA. This training option included a self-paced learning component, instructor-led workshops, and skills coaching circles. Other counties chose to be trained by independent MI trainers. Another group of counties whose workers had been trained recently outside of this pilot were eligible to demonstrate their MI proficiency without further training.

To demonstrate their MI proficiency, all CWPs completed skills demonstrations where they recorded themselves responding to a simulated client encounter. These consisted of 17 submitted learner videos in response to the same set of 15 simulated client videos depicting a mother whose two children had been removed from her care due to substance use issues, speaking directly to the learner. Those who achieved a score of 2 (using MI "to some extent") or above on the Behavior Change Counseling Index (BECCI, Lane et al., 2005), an 11-domain global scale of behavior change counseling, resulting in a total score from 0-4, with higher scores indicating better performance, were certified to claim for MI services in Minnesota. CWPs who attended the training through CPT and MCWTA completed a skills demonstration both before and after training, while those who were trained by an independent trainer before or during the pilot period only completed a skills demonstration after completing their training experience.

Two raters (one with a doctoral degree, a member of the Motivational Interviewing Network of Trainers (MINT), and over 40 years of clinical experience, though not as a CWP; one with a master's degree, accepted to become a member of MINT, and twelve years of clinical experience, but not as a CWP) first coded 153 CWPs' skills demonstrations using the BECCI. These client and learner response videos were edited together to approximate client and child welfare professional "conversations" and then coded by Lyssn (Imel et al., 2019; Imel et al., 2024), an artificial intelligence platform which coded MI skill and calculated a global MI Proficiency score based on the Motivational Interviewing Treatment Integrity scale (MITI, v3.1, Moyers et al., 2016) and Motivational Interviewing Skills Code (MISC, Moyers et al., 2003). The MI Proficiency score was a 0-12 global scale, with higher scores indicating better MI performance, based on other metrics such as CWP empathy, collaboration, percent reflections, percent MI-adherent behaviors, and ratio of reflections-to-questions. Each of the six metrics was scored 0, 1, or 2 based on whether the score met criteria for basic or advanced competency, using cutoffs from the MITI v.3.1.1 manual.

To examine the relationship between human coding with the BECCI and AI coding with the MI Proficiency score, Spearman correlations were used to account for variables that violated the assumption of normality. Moreover, outliers (outside 1.5 times the interquartile range) were removed prior to analyses. Bonferroni correction was used to account for multiple comparisons and significance was set to $p<0.003$. To examine the change in MI skill by coding method, paired samples t-tests and simple linear regression were used (significance was set at $p<0.05$). Receiver operating curve (ROC) analysis was done to explore potentially useful MI Proficiency cutpoint scores that would predict scores of a 2 or above (e.g., certification) on the BECCI. Chi square tests of independence were used to test for significant differences between areas under the curve (AUC) associated with cutpoint scores (Bonferroni corrected significance to $p<0.006$).

Results

Correlations between coding methods

Human-coded BECCI score was significantly correlated with numerous AI-coded indicators of MI skill (Figure 1). First, BECCI score showed significant positive correlations with MI Proficiency score, MI spirit, reflection-to-question ratio, percent MI-adherent behaviors, percent open questions, percent complex reflections, number of times the CWP emphasized control, and percent of time the CWP engaged in active listening. Although BECCI score was positively correlated with empathy score, it was not significant after correcting for multiple comparisons. BECCI score was not significantly correlated with the number of affirmations by a CWP. BECCI score was significantly and negatively correlated with percent MI non-adherent behaviors, especially percent of time the CWP advised and gave information to the client.

Change in MI skill by coding method

From pre to post training, mean human-coded BECCI score increased from 1.6 to 3.0 ($t(35) = 11.0$, $p<0.0001$), and mean AI-coded MI Proficiency score increased from 5.9 to 7.8 ($t(35) = 4.8$, $p<0.0001$) (Figure 2). A simple linear regression was calculated to predict the amount of change in human-coded BECCI score from pre to post training as a function of the amount of change in AI-coded MI Proficiency score. A statistically significant association was found ($\beta = 0.12$, $p<0.05$).

Cutoff scores and coding method

The area under the ROC curve was 0.8 (95% CI, 0.74-0.88), indicating that the overall model could be considered to have good classification utility. An AI-coded MI Proficiency cutpoint of 6 or above, which was associated with a 93.5% sensitivity and a 52.5% specificity, correctly classified the highest percentage (77.1%) of respondents and had the largest area under the curve (0.73) (Figure 3). Chi square tests of independence found that the AUC associated with a cutpoint of 6 was significantly different from all other cutpoints ($p<0.006$), except for 7, 8, and 9.

Conclusions

The present brief sought to compare human-coding of MI skill using the BECCI with AI-coding using an MI Proficiency score rooted in the MITI and MISC. Overall, BECCI scores coded by humans and AI-coded overall MI Proficiency scores were moderately positively correlated. The AI-coded percentage of MI adherent behaviors, percentage of complex reflections, count of time emphasizing control, percentage of time actively listening, and percentage of time giving information were the most strongly correlated with human-coded BECCI score. Moreover, change in human-coded BECCI was associated with change in AI-coded MI Proficiency for pre to post training, and a score of 6 or above on the AI-coded MI Proficiency score appeared best associated with a cutoff score of 2 or above on the human-coded BECCI. Further study could examine what accounts for the differences between these coding systems (e.g., BECCI vs. MITI/MISC, human coders vs. AI-coding). Researchers could also study for whom (e.g., strong performers vs. less strong performers) the two systems best aligned.

Main Findings

Figure 1. Human-coded BECCI score was correlated with AI-coded indicators of MI skill

Human-coded BECCI score correlated with	Spearman's rho *p<0.003
AI-coded:	
MI Proficiency score	0.47*
MI Spirit score	0.31*
Empathy score	0.21
Reflection to question ratio	0.28*
% MI adherent behaviors	0.48*
% open questions	0.41*
% complex reflections	0.63*
Count of times emphasizing control	0.59*
Count of affirmations	0.05
% of time active listening	0.74*
% of time advising	-0.42*
% of time giving information	-0.66*

Figure 2. Change in human-coded BECCI was associated with change in AI-coded MI Proficiency

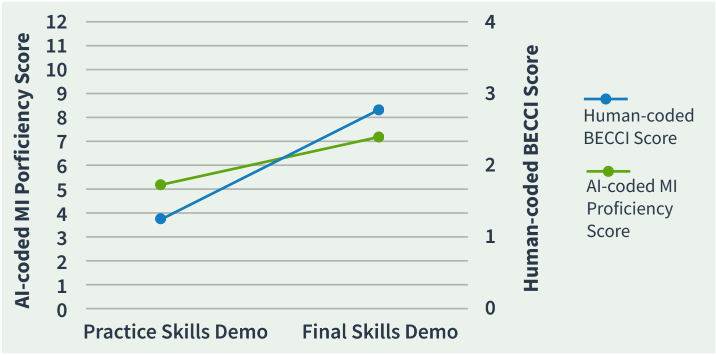

Figure 3. 6 or above on the AI-coded MI Proficiency score appeared best associated with a cutoff score of 2 or above on the human-coded BECCI

AI-coded MI Proficiency Score Cutpoint	Sensitivity (% of BECCI >=2 correctly classified)	Specificity (% of BECCI <2 correctly classified)	Correctly identified	AUROC	95% CI
2>=	100.0%	0.0%	60.1%	—	
3>=	100.0%	3.3%	61.4%	0.52	0.49–0.54
4>=	100.0%	6.6%	62.8%	0.53	0.50-0.56
5>=	97.8%	21.3%	67.3%	0.60	0.54-0.65
6>=	93.5%	52.5%	77.1%	0.73	0.66-0.80
7>=	73.9%	68.9%	71.9%	0.71	0.64-0.79
8>=	54.4%	85.3%	66.7%	0.70	0.63-0.77
9>=	35.9%	93.4%	58.2%	0.65	0.59-0.71
10>=	22.8%	98.4%	52.9%	0.61	0.56-0.65
11>=	10.9%	100.0%	46.4%	0.55	0.52-0.59
12>=	5.4%	100.0%	43.1%	0.53	0.50-0.55

If you would like access to the full list of references for this brief, please contact cptresearch@umn.edu.

SUGGESTED CITATION

Van Wert, M. & Gus, E. (2025). Research Brief: Comparing human and artificial intelligence coding of Motivational Interviewing skills among child welfare professionals (July, 2025). Center for Practice Transformation, University of Minnesota.

REFERENCES

Miller WR. The evolution of motivational interviewing. Behavioural and Cognitive Psychotherapy. 2023;51(6):616-632.

Lane C, Huws-Thomas M, Hood K, Rollnick S, Edwards K, Robling M. Measuring adaptations of motivational interviewing: the development and validation of the behavior change counseling index (BECCI). Patient Educ Couns. 2005 Feb;56(2):166-73.

Cheffer N, Barnett E, Bullara A. Effectiveness of asynchronous motivational interviewing training for pediatric nurse practitioner students: A mixed-methods evaluation. Nurse Educ Today. 2025 Oct;153:106790.

Darnell DA, Parker LE, Engstrom A, Fisher D, Diteman KD, Dunn C. Pragmatic Quality Assessment of Brief Health Behavior Change Interventions: Evidence for Criterion-Related Validity of the Behavior Change Counseling Index. J Clin Med Res. 2019 Nov;11(11):764-768.

Gill I, Oster C, Lawn S. Assessing competence in health professionals' use of motivational interviewing: A systematic review of training and supervision tools. Patient Educ Couns. 2020 Mar;103(3):473-483.

Hershberger PJ, Pei Y, Bricker DA, Crawford TN, Shivakumar A, Castle A, Conway K, Medaramitta R, Rechtin M, Wilson JF. Motivational interviewing skills practice enhanced with artificial intelligence: ReadMI. BMC Med Educ. 2024 Mar 5;24(1):237.

Madson MB, Campbell TC. Measures of fidelity in motivational enhancement: a systematic review. J Subst Abuse Treat. 2006 Jul;31(1):67-73.

Forrester, D., McCambridge, J., Waissbein, C., Emlyn-Jones, R., & Rollnick, S. (2008). Child risk and parental resistance: Can motivational interviewing improve the practice of child and family social workers in working with parental alcohol misuse?. British Journal of Social Work, 38(7), 1302–1319.

Carroll, K. M., Libby, B., Sheehan, J., & Hyland, N. (2001). Motivational interviewing to enhance treatment initiation in substance abusers: An effectiveness study. The American Journal on Addictions, 10(4), 335–339.

Forrester, D., Westlake, D., Killian, M., Antonopoulou, V., McCann, M., Thurnham, A., ... & Hutchison, D. (2018). A randomized controlled trial of training in Motivational Interviewing for child protection. Children and Youth Services Review, 88, 180-190

Chaffin, M., Valle, L. A., Funderburk, B., Gurwitch, R., Silovsky, J., Bard, D., McCoy, C., & Kees, M. (2009). A motivational intervention can improve retention in PCIT for low-motivation child welfare clients. Child Maltreatment, 14(4), 356-368.

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(2), 177–184.

Moyers TB, Rowell LN, Manuel JK, Ernst D, Houck JM. The Motivational Interviewing Treatment Integrity Code (MITI 4): Rationale, Preliminary Reliability and Validity. J Subst Abuse Treat. 2016 Jun;65:36-42.

Imel ZE, Pace BT, Soma CS, Tanana M, Hirsch T, Gibson J, Georgiou P, Narayanan S, Atkins DC. Design feasibility of an automated, machine-learning based feedback system for motivational interviewing. Psychotherapy (Chic). 2019 Jun;56(2):318-328.

Imel ZE, Tanana MJ, Soma CS, et al. Outcomes in Mental Health Counseling From Conversational Content With Transformer-Based Machine Learning. JAMA Netw Open. 2024;7(1):e2352590.