Original Article

Korean J Health Promot 2026;26(3):83-93
pISSN: 2234-2141 • eISSN: 2093-5676
https://doi.org/10.15384/kjhp.2026.00304

Korean Journal of Health Promotion
KJHP

A Preliminary Study of Motivational Interviewing Training Using a Large Language Model (ChatGPT): Effects on Empathy and Motivational Interviewing Confidence among Social Work Students

Ho Yeob KANG, PhD[1], Sohyun WOO, PhD[2], and Hee Jung KIM, PhD[3]

[1]Department of Social Welfare, Dong-A University of Health, Yeongam, Korea
[2]The Research Institute of Nursing Science, College of Nursing, Seoul National University, Seoul, Korea
[3]Department of Nursing, Gachon University, Incheon, Korea

ABSTRACT

Background: Motivational interviewing (MI) requires repeated practice, individualized feedback, and ongoing coaching that short-term group education cannot sufficiently provide. Large language models (LLMs) such as ChatGPT may offer realistic conversational simulation and immediate feedback. This pilot study examined whether an LLM (ChatGPT)–integrated MI training program improves empathy and MI Confidence among social work students.

Methods: A one-group repeated measures quasi-experimental design was applied to 15 undergraduate social welfare students. The four-session program used ChatGPT (GPT-5.5) as a virtual client for open-ended questions, affirmations, reflective listening, and summaries (OARS) practice with automated Motivational Interviewing Treatment Integrity-based feedback. Empathy, MI Skill Confidence, and Global Interviewing Confidence were measured at pretest, posttest, and 4-week follow-up, and analyzed using Friedman and Wilcoxon signed-rank tests. Qualitative content analysis of participants' experiences was also conducted.

Results: Empathy improved significantly across time points (χ^2=6.037, P=0.049), increasing from pretest to posttest (P=0.014) and remaining higher at follow-up (P=0.037). MI Skill Confidence showed an upward trend without statistical significance (χ^2=2.561, P=0.278). Global Interviewing Confidence changed significantly (χ^2=6.682, P=0.035), being higher at follow-up than pretest (P=0.021). Qualitative analysis yielded four themes: realistic simulation, immediate feedback, skill mastery through repeated practice, and requests for enhanced realism.

Conclusions: An LLM (ChatGPT)–integrated MI training program may improve empathy and overall interviewing confidence and function as an effective auxiliary tool complementing traditional MI education. LLMs can support learners' repeated practice and immediate feedback.

Keywords: Motivational interviewing, Empathy, Artificial intelligence, Self efficacy, Simulation training

Received: July 14, 2026; **Revised:** July 31, 2026; **Accepted:** August 10, 2026

Corresponding author: Hee Jung KIM, PhD
Department of Nursing, Gachon University, 191 Hambangmoe-ro, Yeonsu-gu, Incheon 21936, Korea
Tel: +82-32-820-4224 Fax: +82-31-750-8859 E-mail: illine@paran.com

© 2026 The Korean Society for Health Promotion and Disease Prevention
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Motivational interviewing (MI), developed by Miller and Rollnick [1], is a client-centered, collaborative, evidence-based approach that helps clients explore and strengthen their own motivation for change. Over the past four decades, MI has demonstrated effectiveness across diverse fields, including addiction, chronic disease management, mental health, and social welfare, and is widely used in settings requiring health behavior change.

In Korea, MI education and training have been actively conducted in the medical, counseling, and social welfare fields. In social work, where empathic communication forms the foundation of professional relationships, prior studies have reported that MI training improves practitioners' empathy and communication self-efficacy [2,3]. On the basis of these effects, MI has been used as a major practice model in social work practice skills [3,4].

MI training has been delivered through various methods, including lecture, role play and video feedback, and reliable tools such as the Motivational Interviewing Treatment Integrity (MITI) 4.2.1 [5] have been developed to evaluate outcomes. However, because MI requires not merely the acquisition of knowledge but strategic communication competence encompassing advanced relationship-building skills and reflective techniques, short-term education alone makes it difficult to maintain proficiency [3,6]. Moreover, for MI training to be sufficiently translated into and sustained as learner competence, repeated practice, individualized feedback, and ongoing coaching are essential [3,7]. Yet practical constraints such as cost, time, and limited expert resources often confine training to short-term group education [6,8], and new training environments that can complement these limitations are needed.

Recent advances in artificial intelligence (AI) have been changing education and training methods across various professional fields. Generative AI in particular has drawn attention as a tool that supports repeated practice and individualized learning by implementing interactive learning environment and automated feedback [8,9]. Among these, large language models (LLMs) such as ChatGPT can provide realistic conversational simulations giving them potential well suited to communication-based training. LLMs can implement diverse clinical scenarios and provide immediate interaction, thereby creating an environment for repeated practice without the continuous involvement of standardized patients or instructors [10].

Studies on counseling and communication training using AI and LLMs have recently been reported. Some studies have suggested that AI-based training using virtual patients can positively affect the improvement of empathy and patient-centered communication competence [11,12]. In addition, Zhu et al. [13] reported that users with prior ChatGPT experience differed from first-time users in their utterance patterns, use of strategic questions, and frequency of change talk. Thus, prior LLM experience may influence MI-based interactions.

Despite this potential, the clinical suitability of LLMs requires careful examination. Basar et al. [14] analyzed the quality of reflections generated by an LLM and found that the ability to perform deep inference about emotion or to strategically promote change talk was limited. This limitation suggests that LLMs may serve as supportive tools for beginning learners' repeated practice rather than as replacements for skilled counselors [15]. Meyer [16] likewise confirmed that MI principles can be reflected in an LLM to a limited extent but noted that empirical verification of its actual educational effects has yet to be established.

However, few studies have examined how LLM-based MI education affects core outcome variables such as learners' MI Confidence and empathy competence. This study therefore sought to verify whether an LLM (ChatGPT)–integrated MI training program improves trainees' MI Confidence and empathy competence, thereby providing evidence on whether LLM-based training can complement traditional MI education as a supportive tool.

METHODS

Study design

This study was a pilot study to explore the effects of an MI training program using an LLM (ChatGPT) among undergraduate social work students, applying a one-group repeated-measures quasi-experimental design.

Participants

Participants were recruited from undergraduate students enrolled in the Department of Social Welfare at D University, located in J Province, who participated in an extracurricular program. To ensure homogeneity of the participants, eligibility was limited to students with no prior educational experience related to MI or to generative AI. Recruitment was conducted through departmental bulletin-board announcements and promotion; after being informed of the study's purpose, students who vol-

https://doi.org/10.15384/kjhp.2026.00304

untarily expressed an intention to participate were selected.

G*Power 3.1.9.4 was used to determine the sample size. Based on a repeated-measures analysis of variance (ANOVA) with an effect size (f) of 0.25, a significance level (α) of 0.05, a power (1–β) of 0.80, and three measurement points (pretest, posttest, and 4-week follow-up), a minimum of 28 participants was required. Considering possible attrition, recruitment targeted 30 participants; however, only 17 were recruited because participation was voluntary. After excluding 2 incomplete responses, 15 participants were included in the final analysis.

Large language model (ChatGPT)-integrated motivational interviewing training program

The program was based on the MI-based communication training program of Kang et al. [3], and the session structure was organized into a total of four sessions based on Kang and Lim [2]. The goals of the program were to improve empathy competence and to enhance MI Confidence, and the educational content was organized around the MI spirit and the core skills (open-ended questions, affirmations, reflective listening, and summaries, OARS) (Table 1).

Content validity of the program was reviewed by two MI experts, and the Content Validity Index (CVI) was calculated using a 5-point Likert scale. Evaluated according to the criterion of Fehring [17] (CVI≥0.80), both the program content and the session structure showed a CVI of 1.00, confirming content validity.

Each session consisted of a lecture (20 minutes), a demonstration (10 minutes), simulation practice (20 minutes), and instructor feedback (10 minutes). Throughout all sessions, ChatGPT (GPT-5.5) served as a virtual client for practicing OARS, responding in real time to participants' utterances and providing automated feedback based on the MITI 4.2.1 [5]. The practice used a single case of "a client who wants to quit smoking but is experiencing difficulty" (Appendix 1). Applying the same case across all sessions reduced the time needed to adapt to it and allowed OARS to be practiced step by step within the limited training time (Table 2). Practice was conducted individually using participants' personal smartphones, tablets, or desktop computers.

Data collection and intervention

To examine the program's effects, empathy competence and MI Confidence were measured at three time points. The pretest was conducted before the program began, the posttest immediately after it ended, and the follow-up 4 weeks later. Qualitative data were also collected through open-ended questions to explore participants' learning experiences with the LLM.

Measurements

Empathy

Empathy was measured using the adult empathy scale developed by Kim and Kim [18]. The scale consists of 32 items rated on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"). The total score ranges from 32 to 160, with higher scores indicating higher levels of empathy. The reliability (Cron-

Table 1. LLM (ChatGPT)-integrated MI training program

Session	Time (min)	Topic		Learning objectives	Methods
1	60	MI foundations	MI foundations	Understand MI spirit and principles	Lecture, demonstration: Experiencing MI dialogue
			Ambivalence & change talk	Identify and respond to change talk (practice double-sided reflection)	Lecture, demonstration, ChatGPT simulation Practice1: Identify and respond to change talk
2	60	MI skills 1	Open-ended questions	Practice open-ended questions	Lecture, demonstration, ChatGPT simulation Practice 2: Open-ended question
			Affirmation	Practice Affirmation	Lecture, demonstration, ChatGPT simulation Practice 3: Affirmation
3	60	MI skills 2	Simple reflection	Practice simple reflections	Lecture, demonstration, ChatGPT simulation Practice 4: Simple Reflections
			Complex reflection	Practice complex reflections	Lecture, demonstration, ChatGPT simulation Practice 5: Complex reflections
4	60	Integration	Summarizing	Practice summarizing	Lecture, demonstration, ChatGPT simulation Practice 6: Summarizing
			Integration and role-play practice	Conduct a brief MI conversation	ChatGPT role-play simulation

Each session consisted of a 20-minute lecture, a 10-minute demonstration, a 20-minute ChatGPT-based simulation practice, and a 10-minute wrap-up and feedback.
LLM, large language model; MI, motivational interviewing.

Table 2. ChatGPT-based simulation practices used in the MI training program

Session	MI skill	Simulation task	ChatGPT role
1	Change talk	Identifying and responding to change talk	Acts as a client who wants to quit smoking but is experiencing difficulties
2	Open-ended questions	Open-ended question practice	Responds naturally to open-ended questions while expressing change talk
	Affirmation	Affirmation practice	Responds to affirmations while discussing smoking cessation experiences
3	Simple reflection	Simple reflection practice	Responds to simple reflections and maintains the conversation
	Complex reflection	Complex reflections practice	Responds to complex reflections and elaborates on change talk
4	Summarizing	Summarizing practice	Provides extended responses for summarizing practice
	Integration	MI role-play practice	Acts as a smoking cessation client and provides overall feedback on MI skills

All practice sessions employed a standardized smoking-cessation scenario involving a client experiencing difficulty quitting smoking. Participants practiced each MI skill five times with ChatGPT and subsequently received automated feedback based on the Motivational Interviewing Treatment Integrity (MITI) coding framework. The following is an example of the prompts used in the practice sessions: [Prompt Example] Enter the prompts below to assume the role of the counselor and conduct five simple reflections to facilitate a conversation about change with the client (Chat GPT). Chat [GPT Instructions] "Assume you are a client who wants to quit smoking but is struggling, and answer my open-ended questions naturally. After conducting the five simple reflections, provide feedback based on the Motivational enhancement counseling MITI coding system regarding whether the reflections successfully led to a conversation about change."
MI, motivational interviewing.

bach's α) at the time of scale development was 0.93, and in this study it was 0.94.

Motivational interviewing Confidence

(1) Motivational interviewing Skill Confidence
Confidence in performing the core MI skills was measured using the Motivational Interviewing Confidence Survey developed by Larson and Martin [19]. The original instrument consists of 24 items, each rated on an 11-point Likert scale ranging from 0 ("strongly disagree") to 10 ("strongly agree"). The total score of the original scale ranges from 0 to 240, with higher scores indicating greater confidence in MI performance. In this study, considering alignment with the program goals, 10 items directly related to the core skills (OARS) were selected. Accordingly, the total score in this study ranged from 0 to 100, with higher scores indicating greater MI Skill Confidence. The reliability in this study was 0.95. To ensure the content validity of the selected items, they were reviewed by two MI experts.

(2) Global Interviewing Confidence
Overall confidence in conducting interviews was assessed using a single-item scale based on a Numerical Rating Scale. The item consisted of the single question, "Overall, how confident are you in conducting an interview with a client?" rated from 1 ("not at all confident") to 10 ("very confident"). Higher scores indicate greater overall confidence in conducting interviews.

Qualitative data collection

To explore learning experiences and perceptions, applicability, and areas for improvement that are difficult to explain with quantitative results alone, qualitative data on participants' experiences using the LLM (ChatGPT) were collected after the program ended. The questions were, "How was it for you to use the LLM (ChatGPT) in this MI training?" and "What were the strengths and limitations (areas for improvement) of using the LLM (ChatGPT) in this MI training?" To clarify the meaning of the responses and to explore experiences in greater depth, additional individual interviews were conducted with 5 participants. The collected data were integrated and used for qualitative content analysis [20].

Data analysis

Quantitative analysis

Quantitative data were analyzed using IBM SPSS Statistics 24.0 (IBM Corp.). The general characteristics of the participants were summarized using frequencies, percentages, means, and standard deviations (SDs). Program effects were evaluated by analyzing changes in empathy competence and MI Confidence measured at the pretest, posttest, and 4-week follow-up. Because the normality assumption was not satisfied, nonparametric tests were applied. Changes across the three time points were analyzed using the Friedman test, and when significant results were identified, the Wilcoxon Signed-Rank Test was conducted as a post hoc analysis. The statistical significance level was set at $P<0.05$.

Qualitative analysis

The collected qualitative data were analyzed using the qualitative content analysis method [20]. During the analysis, two researchers

 https://doi.org/10.15384/kjhp.2026.00304

independently and repeatedly read the participants' responses, derived meaning units, and conducted first-round coding. Subsequently, through mutual comparison and validity review between the researchers, similar content was categorized and refined, and ultimately four themes and five subthemes were derived.

To ensure the trustworthiness of the analysis, four criteria proposed by Guba [21] were considered: credibility, dependability, transferability, and confirmability. Differences in interpretation between the researchers were reconciled through discussion to reach consensus. The analysis and categorization procedures were systematically recorded using Microsoft Excel to ensure consistency, and the participants, study procedures, and data analysis process were described in detail to enhance transferability.

Ethical consideration

This study was conducted after receiving approval from the Institutional Review Board of Gachon University (approval no. 1044396-202603-HR-036-001). After the study's purpose and procedures were explained, voluntary written consent was obtained from participants, who were informed that they could withdraw from the study at any time without disadvantage. The collected data were anonymized and used only for research purposes.

RESULTS

General characteristics of participants

The general characteristics of the participants are presented in Table 3. There were 15 participants in total: 14 females (93.3%) and 1 male (6.7%). Regarding age, those in their 40s were most common at 8 participants (53.3%), followed by those aged 50 or older at 5 (33.3%) and those in their 30s at 2 (13.3%). The average daily use of generative AI was 10 minutes or more for 8 participants (53.3%) and 5 minutes or less for 7 (46.7%). For familiarity with AI, "moderate" and "familiar" were each most common at 6 participants (40.0%), followed by "very familiar" at 2 (13.3%) and "not at all familiar" at 1 (6.7%).

Changes in outcome variables across measurement points

Empathy

A Friedman test conducted to analyze changes in empathy competence across time points showed a statistically significant difference according to measurement point (χ^2=6.037, P=0.049) (Table 4). A Wilcoxon signed-rank test conducted as a post hoc analysis showed that empathy competence significantly increased from pretest (mean±SD, 133.0±11.88) to posttest (141.6±15.02) (z=–2.449, P=0.014), and the level at the 4-week follow-up (140.4±14.82) also remained significantly higher than at pretest (z=–2.080, P=0.037).

Motivational interviewing Confidence

(1) Motivational interviewing Skill Confidence

A Friedman test conducted to analyze changes in MI Skill Confidence across time points showed no statistically significant difference according to measurement point (χ^2=2.561, P=0.278). The mean scores tended to increase from 69.60±12.34 at pretest to 78.26±9.84 at posttest and 74.93±8.16 at follow-up, but the difference was not statistically significant (Table 5).

Table 3. General characteristics of participants (n=15)

Characteristics	Categories	Number (%)
Sex	Male	1 (6.7)
	Female	14 (93.3)
Age (yr)	30–39	2 (13.3)
	40–49	8 (53.3)
	≥50	5 (33.3)
Daily use of generative AI (ChatGPT) (min)	≤5	7 (46.7)
	≥10	8 (53.3)
For familiarity with AI (ChatGPT)	Not at all familiar	1 (6.7)
	Moderate	6 (40.0)
	Familiar	6 (40.0)
	Very familiar	2 (13.3)

AI, artificial intelligence.

Table 4. Changes in empathy across measurement points (n=15)

Variable	Statistics	Pre	Post	Follow-up	χ^2	P	Post-hoc (Wilcoxon signed-rank test)
Empathy	Mean±SD	133.0±11.88	141.6±15.02	140.4±14.82	6.037	0.049*	Pre-Post: z=–2.449, P=0.014*
	Median (range)	133 (109–153)	141 (117–160)	144 (117–160)			Pre-Follow-up: z=–2.080, P=0.037*
							Post-Follow-up: z=–0.525, P=0.600

Follow-up assessment was conducted 4 weeks after program completion. Friedman test was used to examine overall differences across three measurement points, followed by Wilcoxon signed-rank tests for pairwise comparisons.
Post, posttest; Pre, pretest; SD, standard deviation.
*P<0.05.

Table 5. Changes in MI Confidence across measurement points (n=15)

Variable	Statistics	Pre	Post	Follow-up	χ^2	*P*	Post-hoc (Wilcoxon signed-rank test)
MI Skill Confidence	Mean±SD	69.60±12.34	78.26±9.84	74.93±8.16	2.561	0.278	
	Median (range)	72.0 (40–86)	79.0 (59–100)	74.0 (63–88)			
Global Interview Confidence	Mean±SD	6.20±1.69	7.13±1.59	7.33±1.11	6.682	0.035*	Pre-Post: z=–1.602, *P*=0.109
	Median (range)	7.0 (3–8)	7.0 (3–10)	7.0 (5–9)			Pre-Follow-up: z=–2.312, *P*=0.021*
							Post-Follow-up: z=–0.493, *P*=0.622

MI Skill Confidence refers to confidence in performing OARS-related motivational interviewing skills. Global Interviewing Confidence refers to overall confidence in conducting client interviews. Follow-up assessment was conducted 4 weeks after program completion. Friedman test was used to examine overall differences across three measurement points. Wilcoxon signed-rank tests were conducted as post-hoc analyses when significant overall differences were identified.
MI, motivational interviewing; OARS, open-ended questions, affirmations, reflective listening, and summaries; Post, posttest; Pre, pretest; SD, standard deviation.
*P<0.05.

Table 6. Qualitative findings on the use of ChatGPT in MI training

Theme	Subtheme
Realistic simulation experience	Realistic interaction with a virtual client
Learning facilitated by immediate feedback	Real-time correction and summarization of the core skills (OARS)
Skill development through repeated practice	Repeated practice in a low-pressure setting
Requests for enhanced realism	Formulaic and limited response patterns
	Requests for responses with nonverbal cues

MI, motivational interviewing; OARS, open-ended questions, affirmations, reflective listening, and summaries.

(2) Global Interviewing Confidence

A Friedman test conducted to analyze changes in Global Interviewing Confidence across time points showed a statistically significant difference according to measurement point (χ^2=6.682, *P*=0.035). The mean scores increased from 6.20±1.69 at pretest to 7.13±1.59 at posttest and 7.33±1.11 at follow-up. Post hoc analysis showed no significant differences for the pretest–posttest (z=–1.602, *P*=0.109) or posttest–follow-up (z=–0.493, *P*=0.622) comparisons, whereas the difference between pretest and the 4-week follow-up was statistically significant (z=–2.312, *P*=0.021) (Table 5).

Qualitative findings on the use of ChatGPT in motivational interviewing training

Analysis of the qualitative data yielded four themes and five subthemes (Table 6).

(1) Realistic simulation experience

Participants reported that training using ChatGPT felt like interacting with an actual client and that its immediate responses to their questions made the practice feel like a real counseling session. They also reported becoming more immersed in this process.

"*Because a response actually came (from ChatGPT), I could think about how to ask questions so that the client could move in an appropriate direction.*" *(Participant 2)*

"*The more I did it, the more immersed I became. I really felt like a counselor, and I could genuinely sense what a counselor's basic attitude is.*" *(Participant 4)*

(2) Learning facilitated by immediate feedback

Participants stated that because ChatGPT gave immediate feedback on their utterances, they could quickly check their own responses. They also reported that they could notice ineffective communication patterns such as giving advice and could establish criteria for choosing more appropriate responses in subsequent counseling.

"*Since I got feedback on my responses right away, I could think a bit more about what kinds of questions to ask.*" *(Participant 2)*

"*During counseling, I unconsciously end up giving advice, but the immediate feedback makes me realize I shouldn't do that next time.*" *(Participant 4)*

(3) Skill development through repeated practice

Participants stated that training using ChatGPT could be conducted without constraints of time and place and that they could practice repeatedly. They also noted that being able to respond after thinking things through eased the pressure of making mis-

 https://doi.org/10.15384/kjhp.2026.00304

takes, making it useful for preliminary practice before meeting actual clients.

"*It was nice because I could take longer to think and still continue the conversation.*" *(Participant 1)*

"*I liked that I could practice anytime and anywhere.*" *(Participant 3)*

"*There was no burden even when I made mistakes. Because I kept receiving feedback, I didn't have to worry about whether it was right or wrong. I think it's a good means of practice before meeting an actual client…*" *(Participant 5)*

(4) Requests for enhanced realism

Participants also reported areas for improvement in training with ChatGPT. They noted that its responses tended to be repetitive and did not sufficiently capture the diverse situations that real clients might present, which limited the practice to some extent.

"*There's a feeling that it keeps skimming the surface. The given prompts tend to circle back to similar answers, going round in circles…*" *(Participant 2)*

"*Actual clients can be a bit more complex and present unexpected situations, but ChatGPT sometimes gives predictable and obvious answers.*" *(Participant 5)*

"*It would be even better if (the client's) emotions were also implemented. In a script, the client's actions and facial expressions, not just their words, are described in parentheses. I wish I could know more about the client's (nonverbal) emotions.*" *(Participant 1)*

DISCUSSION

This study examined the effects of an LLM (ChatGPT)–based MI training program on undergraduate social welfare students' empathy competence and MI Confidence. Empathy competence and Global Interviewing Confidence significantly improved after the program, with some effects maintained one month later, and the qualitative analysis identified realistic simulation, immediate feedback, and repeated practice as positive experiences. These results suggest the potential of LLM-based training as a supportive tool in MI education.

Empathy competence significantly improved at both the posttest and the 4-week follow-up, consistent with prior studies reporting positive effects of MI training on empathy [2,3]. Empathy is a core element of the MI spirit and the foundation of the therapeutic relationship [1]. Participants' reports of "realistic counseling experiences" and "immersion" suggest that repeatedly practicing core skills such as reflection and affirmation in realistic counseling situations contributed to this improvement, in line with prior studies on generative AI–based virtual patient training [11,12]. The maintenance of the improvement at one month is also meaningful, given repeated reports that skills are difficult to maintain after short-term workshops [6]. The ease of repeated practice in LLM-based environments may have contributed to this maintenance.

For MI Skill Confidence, an upward tendency in mean scores was observed, but the difference was not statistically significant. This can be interpreted in several respects. First, the participants in this study were beginning counselors; once trained, they may come to clearly recognize their own shortcomings, leading to more conservative self-evaluation [22]. Such recalibration of self-evaluation has been repeatedly reported in the field of medical education as well [23], and may be one possibility explaining why confidence was not statistically significant in this study. Meanwhile, some prior studies have reported significant improvements in MI Confidence after education and training; these commonly introduced training methods that provided learners with intensive, individualized feedback and coaching, such as small-group training and practice with real or standardized patients [24,25]. That is, the training methods commonly suggested by prior studies were 'systematic feedback' and 'ongoing supervision' [7,26]. Future research needs to design programs integrating these elements within a ChatGPT-based environment. The small sample size may also have contributed.

In contrast, Global Interviewing Confidence, measured with a single item, was found to have significantly improved over time. Looking at the related qualitative data, the participants felt less pressure about mistakes, could take time to think, and benefited from repeated practice. These experiences suggest that, although the participants' confidence in each skill was still limited from a beginner's perspective, their overall confidence in counseling increased as they felt less pressure about counseling itself and had the opportunity to review their own approach. Confidence is an important predictor of behavior change; it has been measured in various ways and has yielded various results in prior studies. Therefore, future research needs to examine, together, more appropriate confidence measurement indicators and training strategies that reflect the characteristics of LLM-based MI training.

The qualitative analysis showed the positive aspects and lim-

itations of LLM-based MI training in greater detail. Participants experienced a sense of realism similar to actual counseling through conversations with the virtual client and perceived the immediate-feedback and repeated-practice environment as strengths. These results are consistent with the characteristics of deliberate practice, which improves performance through repeated practice and immediate feedback [27,28]. Immediate and specific feedback is also known to be a key element that enhances learning [29], suggesting that ChatGPT can support a self-directed learning environment by providing such feedback in real time.

At the same time, however, participants also experienced limitations, such as ChatGPT's responses being formulaic and repetitive. As Basar et al. [14] pointed out, this shows that current LLMs perform reasonably well at the level of simple reflection but may have limitations in complex emotional inference or the strategic elicitation of change talk. Therefore, at the present stage, it would be more appropriate to use LLM-based MI training as a supplementary learning tool that supports novice learners' repeated practice and early skill acquisition rather than as a complete replacement for skilled supervisors or actual clinical training. Participants also requested that the virtual client's nonverbal cues and a variety of cases be incorporated into the responses for more realistic practice. This suggests that prompt engineering is a key factor influencing the educational effects of LLM-based counseling training. Future programs will therefore need to develop more advanced prompts that include emotional and nonverbal cues and diverse response scenarios.

The limitations of this study are as follows. First, a one-group quasi-experimental design was applied to 15 students at a single university, which limits both generalization and causal inference. Multicenter randomized controlled trials comparing traditional education with AI-integrated education in adequately powered samples are needed. Second, effects were assessed only with self-report measures, so objective change in interviewing performance was not captured. Future studies should apply an objective structured clinical examination with standardized patients, or have trained raters code recorded interviews using the MITI system. Third, because follow-up was limited to 4 weeks, whether the improved competence is sustained in practice remains unclear. Extending follow-up to 3 and 6 months would verify the durability of the effects and clarify the need for refresher training. Fourth, LLMs are continuously updated, so the same prompt may yield different responses later. Prompts also need refinement to reflect clients' nonverbal cues and diverse cases.

Despite these limitations, this study is meaningful in that it confirmed the applicability and initial effects of an LLM (ChatGPT)-based MI training program, suggested that generative AI can serve as a supportive tool for repeated learning and immediate feedback rather than a replacement for instructors, and provides foundational data for the development of AI-based MI education programs and follow-up research.

ORCID

Ho Yeob KANG, https://orcid.org/0009-0008-1027-8141
Sohyun WOO, https://orcid.org/0000-0002-3198-8854
Hee Jung KIM, https://orcid.org/0000-0001-6555-1813

AUTHOR CONTRIBUTIONS

Dr. Hee Jung KIM had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors reviewed this manuscript and agreed to individual contributions.

Conceptualization: HYK. Data curation: all authors. Formal analysis: all authors. Investigation: all authors. Methodology: HYK. Project administration: HYK. Resources: HYK and HJK. Software: HYK and HJK. Supervision: HYK. Validation: HYK and HJK. Visualization: HYK and HJK. Writing–original draft: SW and HJK. Writing–review & editing: SW and HJK.

CONFLICTS OF INTEREST

No existing or potential conflict of interest relevant to this article was reported.

FUNDING

None.

DATA AVAILABILITY

The dataset supporting the conclusions is available from the corresponding author on reasonable request.

REFERENCES

1. Miller WR, Rollnick S. Motivational interviewing: helping people change and grow. 4th ed. Guilford Press; 2023.

2. Kang HY, Lim SC. Development and evaluation of an integrated motivational interviewing-cognitive behavioral therapy counseling training program for suicide prevention practitioners. J Soc Work Couns 2025;9(1):139-61.
3. Kang HY, Kim HJ, Lim SC. Effects of motivational interviewing-based communication training program for enhance counseling competency of social work undergraduate students. Korean J Soc Welf Educ 2023;61:1-22.
4. Korean Council on Social Welfare Education. 2022 Social welfare curriculum guidelines [Internet]. Korean Council on Social Welfare Education; 2022 [cited 2026 Jul 2]. Available from: http://kcswe.kr/bbs/board.php?bo_table=b0401&wr_id=36
5. Moyers TB, Manuel JK, Ernst D. Motivational Interviewing Treatment Integrity coding manual 4.2.1. University of New Mexico; 2014.
6. Boyle S, Vseteckova J, Higgins M. Impact of motivational interviewing by social workers on service users: a systematic review. Res Soc Work Pract 2019;29(8):863-75.
7. Miller WR, Sorensen JL, Selzer JA, Brigham GS. Disseminating evidence-based practices in substance abuse treatment: a review with suggestions. J Subst Abuse Treat 2006;31(1):25-39.
8. Hershberger PJ, Pei Y, Bricker DA, Crawford TN, Shivakumar A, Vasoya M, et al. Advancing motivational interviewing training with artificial intelligence: ReadMI. Adv Med Educ Pract 2021;12:613-8.
9. Suárez-García RX, Chavez-Castañeda Q, Orrico-Pérez R, Valencia-Marin S, Castañeda-Ramírez AE, Quiñones-Lara E, et al. DIALOGUE: a generative AI-based pre-post simulation study to enhance diagnostic communication in medical students through virtual type 2 diabetes scenarios. Eur J Investig Health Psychol Educ 2025;15(8):152.
10. Chokkakula S, Chong S, Yang B, Jiang H, Yu J, Han R, et al. Quantum leap in medical mentorship: exploring ChatGPT's transition from textbooks to terabytes. Front Med (Lausanne) 2025;12:1517981.
11. Hong H, Shin S. Artificial intelligence-based empathy and compassion training in medical education: impact on patient-centered communication skill development. J Med Life Sci 2025;22(3):81-91.
12. Gilbert A, Carnell S, Lok B, Miles A. Using virtual patients to support empathy training in health care education: an exploratory study. Simul Healthc 2024;19(3):151-7.
13. Zhu J, Dong A, Wang C, Veldhuizen S, Abdelwahab M, Brown A, et al. The impact of ChatGPT exposure on user interactions with a motivational interviewing chatbot: quasi-experimental study. JMIR Form Res 2025;9:e56973.
14. Basar E, Hendrickx I, Krahmer E, Bruijn GJ, Bosse T. To what extent are large language models capable of generating substantial reflections for motivational interviewing counseling chatbots? A human evaluation. Proceedings of the 1st Human-Centered Large Language Modeling Workshop; 2024 Aug 15; Bangkok, Thailand. Association for Computational Linguistics; 2024. p. 41-52.
15. Abid A, Baxter SL. Breaking barriers in behavioral change: the potential of artificial intelligence-driven motivational interviewing. J Glaucoma 2024;33(7):473-7.
16. Meyer S. Integrating motivational interviewing principles and large language models in automated behaviour change support [dissertation]. University of Regensburg; 2025. English.
17. Fehring RJ. Methods to validate nursing diagnoses. Heart Lung 1987;16:625-9.
18. Kim Y, Kim J. Development and validation of empathy scale. Korea J Couns 2017;18(5):61-84.
19. Larson E, Martin BA. Measuring motivational interviewing self-efficacy of pre-service students completing a competency-based motivational interviewing course. Explor Res Clin Soc Pharm 2021;1:100009.
20. Downe-Wamboldt B. Content analysis: method, applications, and issues. Health Care Women Int 1992;13(3):313-21.
21. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Commun Technol J 1981;29(2):75-91.
22. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med 2005;80(10 Suppl):S46-54.
23. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 2006;296:1094-102.
24. Bell K, Cole BA. Improving medical students' success in promoting health behavior change: a curriculum evaluation. J Gen Intern Med 2008;23(9):1503-6.
25. Black B, Lucarelli J, Ingman M, Briskey C. Changes in physical therapist students' self-efficacy for physical activity counseling following a motivational interviewing learning module. J Phys Ther Educ 2016;30(3):28-32.
26. Howard LM, Williams BA. A focused ethnography of baccalaureate nursing students who are using motivational interviewing. J Nurs Scholarsh 2016;48(5):472-81.
27. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev

1993;100(3):363-406.
28. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004;79(10 Suppl):S70-81.
29. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007;77(1):81-112.

 https://doi.org/10.15384/kjhp.2026.00304

Appendix 1. Prompts used in the ChatGPT-based simulation practice

Session	Simulation task	Prompt entered into ChatGPT
1	Identifying and responding to change talk	Take the role of a client who wants to quit smoking but has been unable to keep it up for various reasons. Respond naturally while I conduct the session. The session will run for about 5 minutes. When the 5 minutes are over, evaluate how effectively change talk was elicited, based on the MITI 4.2.1 coding system. Distinguish change talk from sustain talk in the evaluation, and give specific points for improvement.
2	Open-ended questions	Take the role of a client who wants to quit smoking but is having difficulty. Respond naturally to five open-ended questions from me. After answering the fifth question, evaluate the quality of the open-ended questions based on the MITI 4.2.1 coding system. Explain how far each question drew out change talk, and give examples of more effective questions.
	Affirmation	Take the role of a client who wants to quit smoking but is having difficulty. Respond naturally to five affirmations from me. After the fifth affirmation, evaluate how far the affirmations promoted change talk, based on the MITI 4.2.1 coding system. State whether each affirmation was appropriate, and give examples of more effective affirmations.
3	Simple reflection	Take the role of a client who wants to quit smoking but is having difficulty. Respond naturally to five simple reflections from me. After the fifth simple reflection, evaluate how far the simple reflections promoted change talk, based on the MITI 4.2.1 coding system. State whether each reflection was appropriate as a simple reflection, and point out what could be improved.
	Complex reflection	Take the role of a client who wants to quit smoking but is having difficulty. Respond naturally to five complex reflections from me. After the fifth complex reflection, evaluate the appropriateness of the complex reflections and the degree to which they elicited change talk, based on the MITI 4.2.1 coding system. State whether each reflection was appropriate as a complex reflection, and give examples of more effective complex reflections.
4	Summarizing	Take the role of a client who wants to quit smoking but is having difficulty. Give relatively long answers so that I have enough material to summarize. When I give a summary, evaluate how well it reflected the client's change talk, based on the MITI 4.2.1 coding system. Focus the feedback on the accuracy of the summary, the coverage of key content, and whether change talk was emphasized.
	Integrated role-play	Take the role of a client who wants to quit smoking but is having difficulty. The counselor will conduct a session of about 5 minutes using the core MI skills of open-ended questions, affirmation, simple reflection, complex reflection, and summarizing. When the 5 minutes are over, evaluate the fidelity of the session based on the MITI 4.2.1 coding system. Give specific feedback on the ratio of open to closed questions, the ratio of simple to complex reflections, the appropriateness of affirmations, the degree to which change talk was elicited, and the session as a whole, together with points for improvement.

MI, motivational interviewing; MITI, Motivational Interviewing Treatment Integrity.