Journal of Clinical and Nursing Research, 2026, Volume 10, Issue 8
https://www.bbwpublisher.com/journal/JCNR
Online ISSN: 2208-3693
Print ISSN: 2208-3685

Construction and Empirical Study of a Kirkpatrick Four-Level Evaluation Model for a Narrative Medicine Clinical Ideological and Political Education Case Database

Haijiao Wang[1], Jianchao Zheng[2], Kai Yang[1], Lanlan Cai[1], Shuping Gao[1]

[1]Xiangyang No.1 People's Hospital, Hubei University of Medicine, Xiangyang 441000, Hubei, China
[2]The Fourth Clinical College, Hubei University of Medicine, Xiangyang 441000, Hubei, China

Abstract: *Objective:* The current implementation of clinical ideological and political education evaluation is confronted with prominent challenges, including an underdeveloped quantitative evaluation system, a paucity of empirical research data, and difficulty in sustaining long-term educational outcomes. To address this dilemma, this study draws on the Kirkpatrick four-level evaluation model to construct an evaluation framework centered on a narrative medicine clinical ideological and political education case database and examines its practical effects, with the aim of providing a reference for the quantitative evaluation of clinical ideological and political education. *Methods*: Because teaching arrangements precluded random assignment, this study formed a non-randomized concurrent control design based on the year of entry into the department. Sixty clinical medicine interns from Xiangyang First People's Hospital, Hubei University of Medicine, in the 2025 academic year were selected as the control group and received conventional ideological and political teaching; 65 clinical medicine interns in the 2026 academic year were selected as the experimental group and received teaching using the narrative medicine case database. Before and after the intervention, assessments were conducted using the Jefferson Scale of Empathy, a professional identity scale, and objective structured clinical examinations, and verification was performed in conjunction with the quality of parallel charts. *Results*: The satisfaction rate at the reaction level in the experimental group was 92.3%, significantly higher than the 73.3% in the control group, with a statistically significant difference between groups ($P < 0.05$). At the level of professional identity, 84% of the interns in the experimental group demonstrated a high level of identification with the medical profession, far exceeding the 63% in the control group ($P < 0.01$). *Conclusion*: The Kirkpatrick four-level model provides a novel perspective for this study. The application of the narrative medicine case database strengthened medical students' empathy and professional identity and facilitated changes in clinical communication behaviors, demonstrating its value for clinical dissemination.

Keywords: Kirkpatrick model; Narrative medicine; Curriculum ideological and political education; Case database; Effect evaluation

Online publication: Aug 31, 2026

1. Introduction

With the proposal of the "Healthy China" initiative, the cultivation of medical ethics and professional conduct has become a priority in medical education and healthcare development [1]. As a frontline position, clinical medical education must emphasize not only the enhancement of students' clinical diagnostic and therapeutic skills but also the nurturing of humanistic literacy, specifically, the "humanistic foundation and empathic capacity" [2]. Against this background, narrative medicine, with its distinctive features of "close reading" and "reflection", is emerging as a vital link in effective physician-patient communication [3]. Currently, many medical universities in China have embarked on building narrative medicine clinical case databases, achieving genuine integration of professional knowledge and value guidance by recreating authentic clinical scenarios [4]. However, a review of the relevant literature published between 2023 and 2025 reveals a tendency in current case database construction to emphasize quantity over evaluation. Although scholars such as Guo Liping have pointed out that the localized implementation of narrative medicine needs to overcome the problem of "going through the motions", and scholars such as Yao Xu have stressed the need to establish a closed-loop quality monitoring system covering the entire process, in practice, most studies remain limited to using a single Likert scale to collect students' satisfaction [5,6]. The limitation of this approach is that it captures only students' immediate subjective feelings, while whether these feelings can translate into behavioral changes and produce sustained educational effects remains largely unclear. Consequently, the educational efficacy of narrative medicine has long remained in a state of being "difficult to quantify and impossible to measure". This study introduces the four-level model proposed by the American scholar Kirkpatrick [7–10] to construct an evaluation framework tailored to the characteristics of a narrative medicine clinical ideological and political education case database. This framework aims to more objectively assess medical students' professional identity and empathy, with the expectation of providing a new pathway and method for the quantitative evaluation of medical humanities education.

2. Research methods and materials

2.1. Study participants and ethical considerations

A non-randomized concurrent controlled design was adopted. Clinical medicine interns from our hospital during the 2025–2026 academic year were selected and grouped according to the year of entry into the department: the 2025 cohort served as the control group (n = 60, receiving conventional teaching), and the 2026 cohort served as the experimental group (n = 65, receiving case database teaching). Baseline comparison indicated that the two groups were well balanced and comparable in terms of age, sex, and prior academic performance ($p > 0.05$). The study was approved by the hospital ethics committee (Approval No. XY2025-LL-048), and all interns signed informed consent forms upon entering the department. Based on a G*Power 3.1 calculation, with α set at 0.05, power at 0.80, and effect size at 0.5, and anticipating a 10% dropout rate, 60 participants per group were required. The actual sample size met this requirement.

2.2. Construction of the narrative medicine clinical ideological and political education case database

The case database was constructed following the principle of "originating from clinical practice and surpassing clinical practice". Drawing on authentic hospital diagnostic and treatment data, cases with typical humanistic conflicts were selected after removing patients' private information. The case database comprises

four modules:

(1) Parallel charts and reflective journals, emphasizing emotional resonance;
(2) Cases involving high-conflict ethical decision-making (e.g., withdrawal of life-sustaining treatment in end-of-life care, allocation of medical resources), emphasizing value clarification;
(3) Multidisciplinary team (MDT) collaborative narratives, emphasizing the cultivation of professionalism;
(4) Physician growth narratives, emphasizing the construction of professional identity. All cases were jointly reviewed by clinical experts and ideological and political education instructors to ensure the implicit integration of ideological and political elements [11].

2.3. Construction of the evaluation system based on the Kirkpatrick model

To ensure the scientific rigor of the evaluation system, this study employed the Delphi method to determine indicator weights. Fifteen experts from the fields of clinical medicine, medical education, ethics, and statistics were selected to participate in two rounds of consultation. The expert authority coefficient (Cr) was 0.82, and Kendall's coefficient of concordance (W) was 0.78 ($p < 0.01$), indicating a high degree of consensus among experts and the credibility of the weight results.

As shown in **Table 1**, when constructing the evaluation system based on the Kirkpatrick model, this study assigned the highest weight of 30% to the behavior level. This was intended to circumvent the risk of "high scores but low competence", as the ultimate goal of ideological and political education in medicine is to bring about changes in clinical practice, rather than mere classroom satisfaction. Only by observing the interns' actual clinical behaviors can the real educational effects be verified, thereby highlighting the orientation toward "the unity of knowledge and action". The evaluation instruments adopted a mixed design: the Jefferson Scale of Empathy (JSPE) and a self-developed professional identity scale were used to measure cognitive internalization, while standardized patients (SPs) in OSCE stations were employed to record the entire process, allowing objective quantification of students' communication behaviors and ethical decision-making [12].

Table 1. Kirkpatrick's four-level model

Level 1 indicator	Level 2 indicator	Level 3 indicator	Evaluation criterion	Evaluation methodology
Reaction level (20%)	Teaching content recognition (7%)	Authenticity and clinical relevance of cases	All selected cases are derived from real medical records of our hospital in the past 3 years, containing at least one typical humanistic conflict point (e.g., patient refusal of treatment, disagreement among family members on decisions).	Independently scored by 2 teachers with more than 5 years of experience in ideological and political education, combined with the mean of questionnaire results from interns after class.
		Naturalness of integrating ideological and political elements	Ideological and political elements are smoothly connected with narrative details and physicianpatient empathy, without forced didactic content.	Inclass observation + anonymous questionnaire, with a trap question "Did you feel forced didactic content?" inserted in the questionnaire.
Learning level (25%)	Mastery of narrative core competencies (8%)	Narrative analysis and deconstruction ability	Able to accurately extract 3 types of core information from patient narratives: main symptoms, emotional needs, and family situation, and identify potential conflict points in physicianpatient communication.	Case analysis conducted, double-masked scoring by clinical instructors and ideological and political teachers.

Behavior level (30%)	Clinical empathy and communication behaviors (11%)	Narrative inquiry and story collection behaviors	Actively pay attention to 3 types of information during historytaking: changes in lifestyle after illness, psychological changes, and family support, rather than focusing solely on symptoms and signs.	Random check of historytaking records + weekly evaluation scores by instructors, summarized monthly.
Results level (25%)	Longterm educational effectiveness and demonstrative impact (8%)	Positive impact on subsequent clinical practice	Within 3 months after internship, still able to proactively include relevant psychological and social assessment items in medical records, with stable humanistic communication behaviors.	Followup visit by instructors 3 months after internship, combined with random inspection of outpatient medical records for evaluation.

3. Results

3.1. Baseline characteristics

A total of 125 interns were enrolled (60 in the control group and 65 in the experimental group). The two groups were well balanced and comparable at baseline: age (22.4 ± 0.8 years vs. 22.6 ± 0.7 years) and sex ratio (male/female: 31/29 vs. 33/32) showed no significant differences ($p < 0.05$).

3.2. Evaluation of intervention effects

As shown in **Table 2**, the experimental group outperformed the control group on all evaluation levels ($p < 0.05$ or $p < 0.001$).

(1) Reaction level

The overall educational satisfaction rate in the experimental group was 92.3%, which was higher than 73.3% in the control group ($t = 7.152$, $p < 0.001$). In the interviews, students reported that the narrative materials in the case database were more emotionally compelling than traditional didactic teaching.

(2) Learning level

The Jefferson Scale of Empathy score in the experimental group was 125.4 ± 8.2, significantly higher than the 108.5 ± 9.1 in the control group ($p < 0.001$). The correct rate on the ethics assessment for "breaking bad news" in the experimental group was 92.0%, an increase of 11 percentage points compared to 81.0% in the control group.

(3) Behavior level

In the OSCE assessment, the number of open-ended questions and empathic responses in the experimental group was 6.2 ± 1.1, significantly higher than the 2.1 ± 0.8 in the control group ($p < 0.001$). In the parallel charts, the number of psychosocial assessment items in the experimental group was 4.5 ± 0.8, significantly greater than the 1.8±0.6 in the control group ($p < 0.001$).

(4) Results level

At the end of the internship, the professional identity (rated as "very willing" to pursue the career) in the experimental group reached 84.0%, significantly higher than the 63.0% in the control group ($p < 0.01$), suggesting that narrative medicine has a long-term, sustained effect on shaping the professional values of medical students.

Table 2. Multidimensional comparative analysis of the empirical study on teaching effectiveness of the case database

Assessment level	Core indicator	Experimental group ($\bar{x}$ ± s/%)	Control group ($\bar{x}$ ± s/%)	*t* value	*p* value
Reaction level	Overall teaching satisfaction	92.3	73.3	7.152	< 0.001
	Jefferson Empathy Scale score	125.4 ± 8.2	108.5 ± 9.1	8.964	< 0.001
	Ethics assessment accuracy (breaking bad news)	92.0	81.0	4.281	0.001
Learning level	Narrative competence composite score	85.6 ± 5.3	76.2 ± 6.1	7.526	< 0.001
	Empathy assessment score	87.1 ± 4.8	77.5 ± 5.6	8.135	< 0.001
	Ideological and political cognition composite score	84.9 ± 6.0	75.8 ± 6.7	6.987	< 0.001
	Achievement rate of patient communication behavior	86.5	68.2	6.325	< 0.001
	Achievement rate of regular reflection behavior	84.1	65.7	6.114	< 0.001
Behavior level	Correct response rate to ethical dilemmas	89.3	70.1	6.892	< 0.001
	Number of psychosocial assessment items in parallel medical records	4.5 ± 0.8	1.8 ± 0.6	9.257	< 0.001
	Number of open-ended questions and emotional responses in OSCE	6.2 ± 1.1	2.1 ± 0.8	11.683	< 0.001
Results level	Professional identity composite score	84.2 ± 6.0	78.1 ± 6.5	5.021	< 0.001
	Humanistic literacy composite score	85.7 ± 5.3	79.2 ± 5.9	5.436	< 0.001
	Clinical comprehensive ability score	83.3 ± 6.4	77.1 ± 7.1	4.712	< 0.001
	Willingness to choose medical career (percentage very willing)	84.0	63.0	3.105	0.003

Note: All indicators were analyzed using the independent samples *t*-test, and a *p*-value < 0.05 was considered statistically significant for between-group differences.

4. Discussion

4.1. The Kirkpatrick model clarifies the mechanism of clinical ideological and political education evaluation

Previous evaluations of ideological and political education in clinical nursing mostly remained at the superficial level of "paper-based tests" and "post-class questionnaires", which could hardly reflect the actual educational outcomes. This study introduced the Kirkpatrick four-level model to construct a full-chain evaluation system. The empirical data showed that the frequency of open-ended questioning during OSCEs in the experimental group reached 6.2 ± 1.1 times, which was 2.95 times that of the control group. This difference precisely explains why some previous ideological and political education programs with high satisfaction rates failed to improve the quality of nurse-patient communication genuinely. The Kirkpatrick model not only covers students' subjective perceptions of the teaching format (Reaction level) but also quantifies actual changes in clinical communication behavior (Behavior level) through tools such as OSCEs and parallel chart analysis. This extends the evaluation of nursing ideological and political education from "subjective perception" to "objective behavior", providing a practical pathway to demystify the "black box" of evaluation.

4.2. The unique value of narrative medicine as a vehicle for nursing ideological and political education

The results of this study showed that the Jefferson Scale of Empathy score (125.4 ± 8.2) and the proportion of professional identity measured by "very willing to engage in clinical nursing work" (85.0%) in the experimental group were significantly higher than those in the control group ($p < 0.001$). These findings corroborate the viewpoint proposed by Yao Xu and others that nursing ideological and political education requires a closed-loop quality control mechanism of "design–implementation–feedback". Further analysis revealed that the advantage of narrative medicine lies in constructing a logical chain of "emotional triggering–value identification–behavior consolidation". Taking the narrative material of an advanced gastric cancer patient refusing chemotherapy included in this case database as an example, 87% of the interns in the experimental group actively discussed "how to enable the patient to die with dignity", a proportion far higher than the 29% in the control group. This embodied cognitive experience derived from patients' illness narratives avoids the "indoctrination feeling" of traditional didactic ideological and political education and serves as a crucial support for enhancing professional identity.

4.3. Practical implications of setting the weight of the behavior level

This study intentionally increased the weight of the Behavior level to 30%, which was higher than that of the Reaction and Results levels. The clinical data validated the rationality of this design: the overall teaching satisfaction rate in the control group reached 72.4%, yet only 31.7% of the interns could proactively include three or more psychosocial assessment items in their parallel charts, demonstrating that high satisfaction does not equate to behavioral improvement. In the experimental group, the number of psychosocial assessment items in parallel charts reached 4.5 ± 0.8 items, directly reflecting the transformation in nursing interns' thinking from "focusing on the disease" to "focusing on the person". This suggests that in subsequent reforms of nursing ideological and political education, direct clinical behavior observation tools such as the Mini-CEX (especially the nursing communication-specific assessment) and OSCEs should be used as the core of evaluation, rather than relying solely on post-course questionnaire feedback.

4.4. Limitations and prospects

This study still has two main limitations. First, due to the constraints of clinical teaching arrangements, a non-randomized concurrent controlled design was adopted. Although baseline corrections were applied, the 2026 teaching team conducted two additional narrative nursing training sessions compared to 2025, which may have introduced a Hawthorne effect. Second, the follow-up period only covered the internship stage, resulting in insufficient observation of long-term indicators such as professional identity. In the future, a multicenter randomized controlled trial is planned in collaboration with the nursing teaching and research departments of three tertiary-level A hospitals in Northwest Hubei Province, and a tracking cohort will be established to monitor professional identity for three years post-graduation to verify further the long-term effects of narrative nursing ideological and political education.

The evaluation system for the narrative nursing ideological and political education case database based on the Kirkpatrick model constructed in this study overcomes the shortcomings of traditional evaluation, which emphasizes "subjective perception over behavioral transformation". Empirical evidence has confirmed that this model can significantly enhance nursing interns' empathy and professional identity ($p < 0.05$), achieving the transformation of educational effectiveness from "cognitive recognition" to "behavioral

implementation". This system is both scientific and practical, making it suitable for promotion and application in clinical teaching bases.

Funding

Hubei University of Medicine Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era Research Center Project (Project No.: HBYYXSX2601)

Disclosure statement

The authors declare no conflict of interest.

References

[1] He X, Han Z, Ding Y, 2019, Theoretical and Practical Significance of Strengthening Ideal and Belief Education for the Construction of Medical Ethics and Conduct. Chinese Medical Ethics, 32(12): 1593–1594.

[2] Li Z, Liu Y, 2024, Current Status and Influencing Factors of Medical Narrative Competence Among Clinical Medicine Graduates in Hebei Province. Journal of Hebei Medical University, 45(12): 1445–1450.

[3] Xu R, Yang J, 2026, Application of Learning-Situation-Oriented Modular Reconstruction of "Literature and Medicine" in the Stratified Cultivation of Narrative Medicine Competence Among Medical Students. Medical Education Research and Practice, 34(2): 286–291.

[4] Hong F, Hu Y, Liu P, 2017, Research Progress and Implications of Narrative Medicine Education. Chinese Journal of Nursing Education, 14(7): 530–534.

[5] Guo L, Zhu L, Huang R, et al., 2023, Chinese Expert Consensus on Narrative Medicine (2023). Narrative Medicine, 6(6): 381–411.

[6] Zhou X, Cao J, Yao X, et al., 2026, Exploration and Evaluation of a Blended Teaching Faculty Training Curriculum System Oriented by Post Competency. Chinese Journal of Graduate Medical Education, 10(1): 44–47.

[7] Kirkpatrick DL, 1959, Techniques for Evaluating Training Programs: Reaction. American Society for Training and Development Journal, 18: 3–9.

[8] Kirkpatrick DL, 1959, Techniques for Evaluating Training Programs: Learning. American Society for Training and Development Journal, 18: 21–26.

[9] Kirkpatrick DL, 1960, Techniques for Evaluating Training Programs: Behavior. American Society for Training and Development Journal, 19: 13–18.

[10] Kirkpatrick DL, 1960, Techniques for Evaluating Training Programs: Results. American Society for Training and Development Journal, 18: 28–32.

[11] Hua Q, Dai X, Yin X, et al., 2022, Exploration on the Cultivation Path of Humanistic Care Ability Among Medical Students from the Perspective of Narrative Medicine. Medicine and Philosophy, 43(13): 68–72.

[12] Chen L, Shen Y, 2025, Effect of Empathy Training on the Learning Outcomes of Ultrasound-Guided Brachial Plexus Block Among Resident Physicians. China Higher Medical Education, 2025(1): 111–112.