Journal of Multidisciplinary Healthcare

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Development of an Evaluation Index System for Chinese Medicine Specialist Nurse Training Effectiveness: A Delphi-AHP Study

Yanhuan Qi[1], Haimin Li[2,3], Jiahui Mao[2,3], Fei Yu[2], Yun Chen[3], Yinqin Zhong[4], Jianrong Lu[1], Yongzhu Gu[3], Wei Zhang[1]

[1]Department of Endocrinology, Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, Shenzhen, People's Republic of China; [2]The Sixth Clinical Medical College of Guangzhou University of Traditional Chinese Medicine, Shenzhen, People's Republic of China; [3]Nursing Department, Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, Shenzhen, Guangdong Province, People's Republic of China; [4]Director's Office, Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, Shenzhen, People's Republic of China

Correspondence: Yinqin Zhong, Director's Office, Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, 6001 North Ring Avenue, Futian District, Shenzhen, Guangdong Province, People's Republic of China, Email yinqinzhong2025@163.com; Yun Chen, Nursing Department, Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, 6001 North Ring Avenue, Futian District, Shenzhen, Guangdong Province, People's Republic of China, Email chenyun15813833533@163.com

Objective: To develop an evaluation index system for Chinese medicine specialist nurse training effectiveness based on the Context, Input, Process, and Product (CIPP)-Kirkpatrick model, and to preliminarily assess its feasibility in practical application.

Methods: From January to September 2024, a research team developed an initial indicator pool through literature reviews and semi-structured interviews using the CIPP and Kirkpatrick models as theoretical frameworks. The Delphi method was employed for two rounds of consultation. 16 experts from 6 provinces were invited, and 15 completed both rounds, to refine the indicators. Indicator weights were determined using the Analytic Hierarchy Process (AHP). The resulting index system was then preliminarily applied at one provincial Chinese medicine training base in Guangdong Province to assess the feasibility of its practical use.

Results: The two rounds of expert consultation yielded effective response rates of 93.75% and 100%, with an expert authority coefficient of 0.965. Kendall's coefficients of concordance were 0.275 and 0.292, respectively ($P < 0.001$). The final system comprises 4 primary, 13 secondary, and 71 tertiary indicators. Among the primary dimensions, outcome evaluation received the highest weight, followed by process, context, and input evaluation. In the empirical application, the training base achieved an overall score of 82.55. The process evaluation dimension scored highest (90.99), while the outcome evaluation dimension scored lowest (75.93), with research-related indicators scoring below 35 points.

Conclusion: The developed index system demonstrated good content relevance and expert consensus, while reflecting the core features of Chinese medicine syndrome differentiation-based nursing. Preliminary field application suggests that the system is feasible for identifying potential weaknesses in training programs and may serve as a practical reference for the evaluation and improvement of Chinese medicine specialist nurse training. It may also provide a contextual reference for evaluating specialist nurse training programs in other culturally specific or discipline-based practice settings.

Keywords: traditional Chinese medicine, specialist nurse, training effectiveness, evaluation system, delphi technique, analytic hierarchy process

Introduction

With the global development of healthcare systems, nursing specialization has become an important strategy for improving care quality and strengthening professional practice. Specialist nurses are generally regarded as clinically experienced nurses who have completed standardized continuing education and obtained specialized certifications.[1] Studies have demonstrated that specialty nurses play a vital role in improving the quality of patient care, promoting professional development, and reducing healthcare costs.[2,3]

https://doi.org/10.2147/JMDH.S601386
Received: 24 February 2026
Accepted: 8 May 2026
Published: 18 May 2026

In China, Traditional Chinese Medicine (TCM) nursing is rooted in a holistic philosophy that emphasizes prevention, individualized care, and rehabilitation. The *Traditional Medicine Strategy (2014–2023)* emphasizes that standardized training of traditional medicine nursing personnel is central to improving service quality.[4] As healthcare needs continue to diversify, the specialization of Chinese medicine nursing has attracted increasing attention. In recent years, China has made significant progress in establishing a cultivation system for TCM specialist nurses, with a focus on developing training bases, instructional methods, and clinical practice. A survey revealed that 20 out of 29 provinces (69.00%) have launched training programs for TCM specialist nurses.[5] However, despite this growth, training content, implementation, and evaluation remain heterogeneous across regions, largely because no nationally standardized evaluation framework has yet been established. Current programs are mainly organized by provincial authorities or nursing associations. While such approaches provide flexibility, they also result in considerable variation in training quality, competency assessment, and outcome measurement.

Guided by national policy documents and professional standards, existing training programs for TCM specialist nurses generally aim to cultivate advanced theoretical knowledge, syndrome differentiation-based nursing competence, proficiency in characteristic Chinese medicine techniques, and basic research capabilities.[6,7] Training programs typically employ a modular design, integrating theoretical instruction with extensive clinical practice, which usually lasts three to six months. Training is conducted in accredited tertiary-level TCM hospitals under the supervision of senior nursing experts who serve as clinical mentors.[8] Although this training model has supported the rapid development of the field, corresponding evaluation practices remain underdeveloped. Current evaluation approaches are often fragmented, rely heavily on end-of-training or single-outcome assessments, and provide limited feedback on the quality of the educational process. In addition, many existing assessments do not adequately capture core Chinese medicine competencies, particularly syndrome differentiation-based nursing and clinical reasoning. As a result, current evaluation practices are insufficient to support quality assurance, program refinement, and standardized certification in Chinese medicine specialist nurse training.

Currently, research on TCM specialist nurses is increasing annually, with most studies focusing on current practices, training needs, educational models, core competencies, and nursing management,[9] rather than on structured evaluation of training effectiveness. More importantly, few studies have developed transparent, theory-informed, and consensus-based frameworks that simultaneously assess training context, implementation process, and outcome achievement. As Singh has argued, all nursing education programs should undergo systematic evaluation to ensure accountability, guide educational planning, and optimize teaching processes.[10] This represents a critical gap; the absence of standardized evaluation criteria constrains the overall quality of training, complicates post-training certification and workforce deployment, and ultimately hinders the development of TCM nursing as a discipline.

Beyond the Chinese and TCM-specific context, these challenges reflect a broader global issue in the development of specialist nursing roles, where the rapid expansion of training programs often outpaces the establishment of standardized evaluation systems. In many healthcare settings, variability in training content, competency assessment, and certification criteria continues to hinder cross-regional comparability and workforce mobility. Therefore, developing a robust and transferable evaluation framework is not only critical for strengthening TCM specialist nurse training in China but also offers valuable insights for the global advancement of standardized, competency-based specialist nursing education.

Theoretical Framework

Among existing models for evaluating educational and training programs, the CIPP model and the Kirkpatrick model are the most widely applied.[11] The CIPP model comprises four components: context, input, process, and product, providing a dynamic and holistic framework for continuous program evaluation and improvement. By contrast, the Kirkpatrick model consists of four levels: reaction, learning, behavior, and results. Each model offers important strengths, but each also has limitations when used alone. The CIPP model is well-suited to full-process monitoring but provides less detailed guidance for outcome-level assessment, whereas the Kirkpatrick model offers clearer outcome categories but pays less attention to contextual and implementation factors. Previous research has validated the scientific merit of this combined approach, suggesting that its integration offers a more balanced and multidimensional evaluation approach.[12–14]

For TCM specialist nurse training, an effective evaluation framework must assess not only whether trainees achieve expected competencies but also whether the training environment, curriculum design, and implementation process

adequately support those outcomes. Therefore, this study adopted an integrated CIPP-Kirkpatrick framework as an a priori structure for developing the indicator system. The added value of this approach lies not merely in model integration, but in its ability to translate a multidimensional evaluation logic into a structured and consensus-informed index system tailored to the characteristics of Chinese medicine specialist nurse training. Such a framework may also provide a useful reference for evaluating specialist nurse education in other culturally specific, competency-based, or discipline-oriented training settings. The characteristics of both models and the final integration strategy used in this study are presented in Table 1.

Based on this synthesis, this study defines training effectiveness not merely as a result but as a comprehensive measure encompassing both the quality of the educational process and the competencies achieved by trainees. Accordingly, this study aimed to develop an evaluation index system for the training effectiveness of TCM specialist nurses based on this integrated CIPP–Kirkpatrick framework. The resulting tool is expected to serve as a practical reference for promoting standardization and quality assurance in TCM specialist nurse training.

Methods

Study Design

This study used a phased, multi-method design to develop an evaluation index system for TCM specialist nurse training effectiveness, integrating a literature review, semi-structured interviews, two rounds of Delphi consultation, and the Analytic Hierarchy Process (AHP). An integrated CIPP-Kirkpatrick framework was specified a priori to guide the conceptual structure of the indicator system before expert consultation. In this framework, the CIPP model informed the four primary evaluation domains, while the Kirkpatrick model was used to refine outcome-related dimensions and to ensure that the framework captured trainee reaction, learning, behavioral transfer, and training results. The Delphi process was then used to refine indicator content, establish expert consensus, and support indicator weighting within this predetermined conceptual structure, rather than to provide full validation of the tool. The overall research process is shown in Figure 1. The study was conducted and reported in accordance with the Conducting and Reporting Delphi Studies (CREDES) guidelines.[15–17] See Supplementary Table S1.

Research Team

The study was led by an internal multidisciplinary research team of seven members with extensive experience in TCM nursing, nursing education, and quantitative research. The team comprised one dean with a nursing background, one director of the nursing department, one nursing officer, two head nurses, one research nurse, and one statistician.

Preliminary Item Generation

An initial pool of candidate indicators was generated in two sequential steps: literature review and semi-structured interviews.

First, a systematic literature review was performed across eight databases (PubMed, Embase, Web of Science, The Cochrane Library, CNKI, Wanfang, VIP, and SinoMed) up to January 1st, 2024. The search strategy combined subject headings and free-text terms related to "TCM nursing", "specialist nurse", "training", "competency", "education evaluation", and "training effectiveness". Figure 2 shows the search strategy of PubMed. Inclusion criteria were: (1) studies addressing TCM nursing education, specialist nurse training, or competency evaluation; (2) empirical studies,

Table 1 Characteristics of the CIPP and Kirkpatrick Models and Their Integration Strategy

Model	Strengths	Limitations	Role in the Integrated Framework
CIPP	Full-process and dynamic evaluation	More descriptive than analytical; provides limited stratification of outcome levels.	Provides the overall evaluation structure
Kirkpatrick	Clear stratification of training outcomes	Limited attention to pre-training context and implementation process	The reaction level of the Kirkpatrick model is embedded in the Process evaluation of the CIPP model, and the learning, behavior, and results level of the Kirkpatrick model are used as Product evaluation of the CIPP model

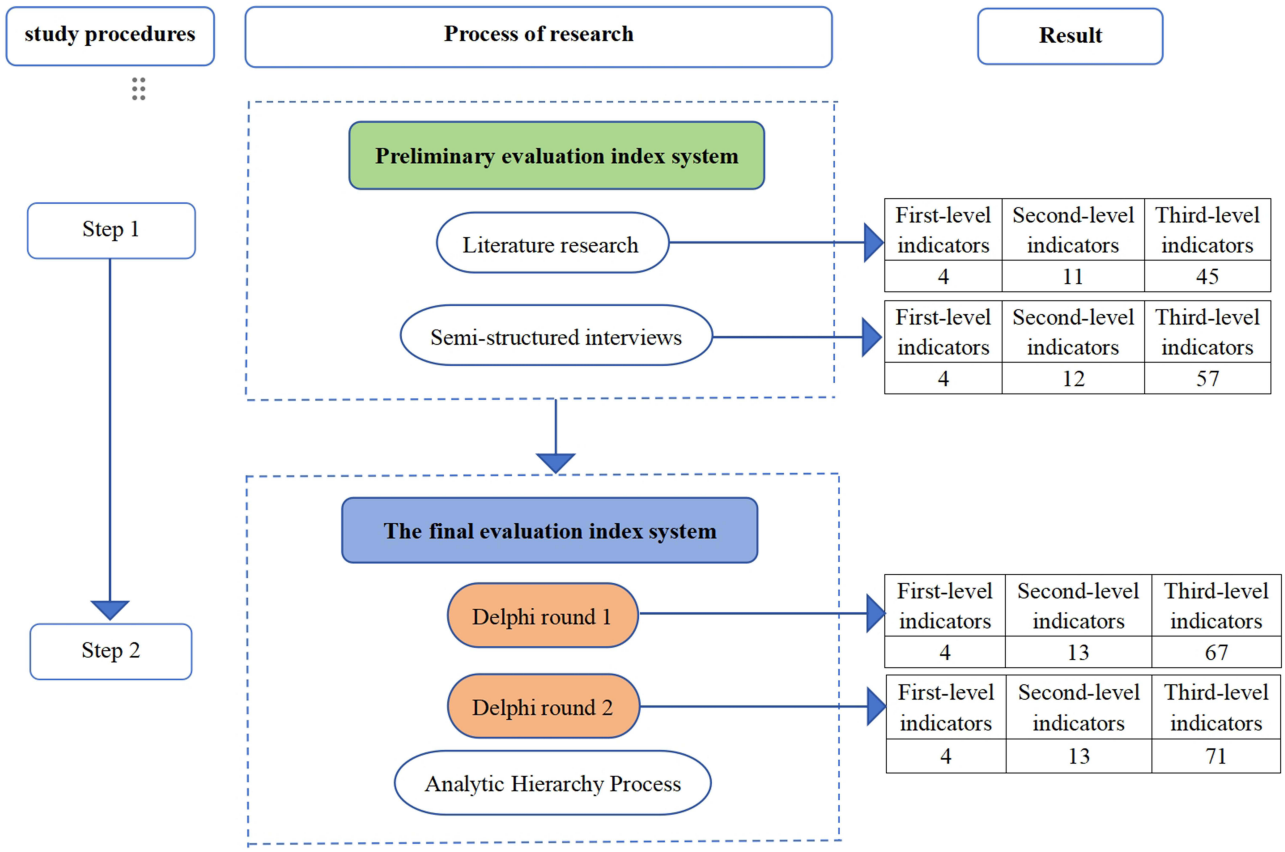

Figure 1 Study procedures.

#1 : (Traditional Chinese Medicine [Title/Abstract]) OR (Traditional Medicine, Chinese [Title/Abstract]) OR (Chinese Traditional Medicine [Title/Abstract]) OR (Medicine, Chinese Traditional [Mesh])

#2: (specialist nurse [Title/Abstract]) OR (specialized nurse [Title/Abstract])

#3: (training [Title/Abstract]) OR (education [MeSH]) OR (((training effect [Title/Abstract]) OR (quality evaluation [Title/Abstract])

#4: (evaluation index [Title/Abstract]) OR (evaluation indicators [Title/Abstract])

#5: #1 AND #2 AND #3 AND #4

Figure 2 PubMed search strategy.

guidelines, consensus statements, or evaluation frameworks; and (3) publications in English or Chinese. Exclusion criteria included: (1) studies unrelated to nursing or TCM practice; (2) commentaries without methodological descriptions; and (3) duplicate publications. Two researchers independently screened titles and abstracts, followed by full-text review. Discrepancies were resolved through discussion or consultation with a third researcher. The process of literature screening is shown in Figure 3. To operationalize this theoretical integration, the four CIPP dimensions were directly mapped to the first-level indicators to form the structural skeleton. Concurrently, Kirkpatrick's levels were used to refine the subdomains: the "Reaction" level was integrated into the CIPP Process evaluation, while the "Learning", "Behavior", and "Results" levels were sequentially embedded within the CIPP Product evaluation. Following these operational rules, extracted content related to training objectives, educational processes, competency evaluation, and training outcomes was

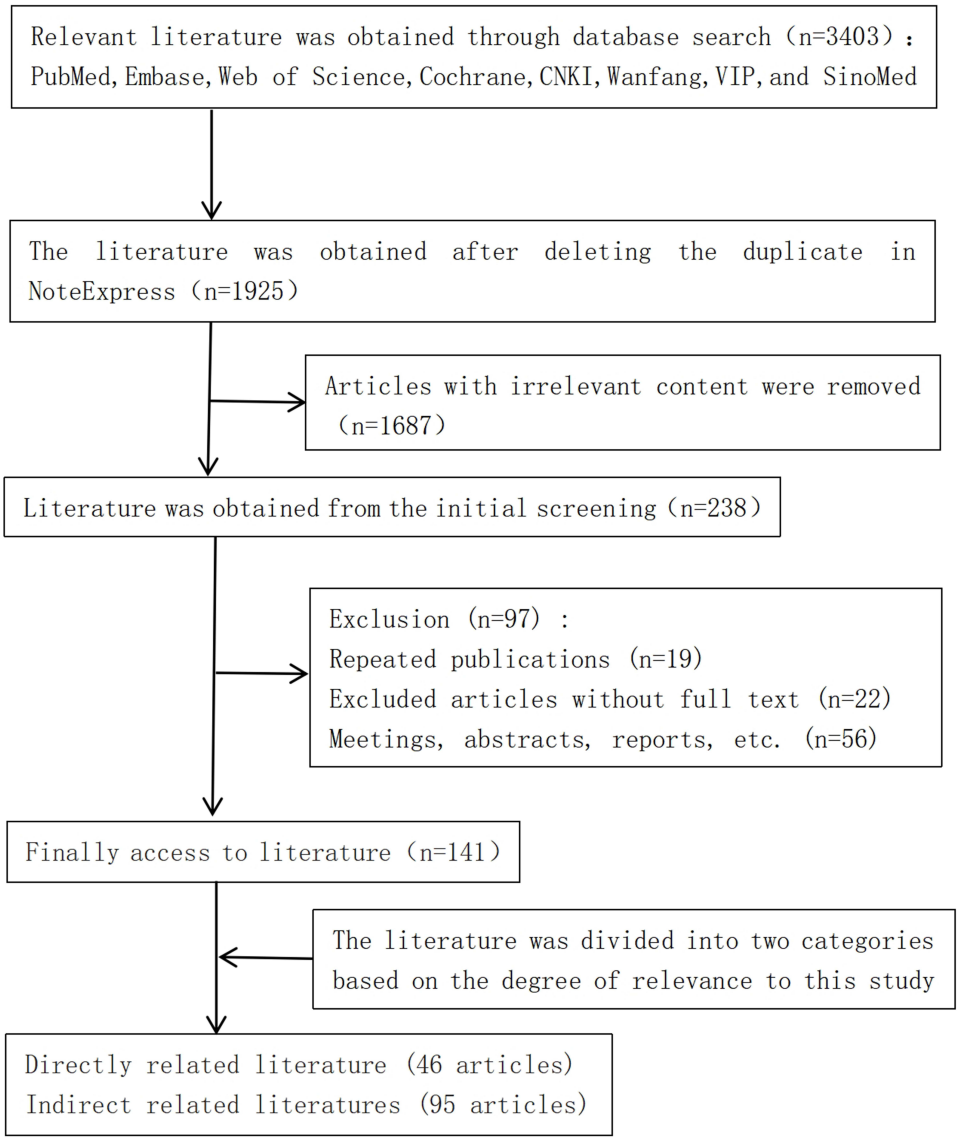

Figure 3 The literature screening process.

coded and mapped accordingly. On this basis, this initial framework consisted of 4 first-level indicators, 12 second-level indicators, and 45 third-level indicators. Theoretical mapping of the initial indicator system is presented in Table S1.1 of Supplementary File 1.

Second, semi-structured interviews were conducted from January to April 2024 to identify indicators that were insufficiently represented in the literature and to assess the practical relevance, clarity, and applicability of the literature-derived items. The research team designed the interview guide based on the results of the literature review. The guide was reviewed by two experts in nursing education, both with extensive experience in qualitative research, to ensure a logical flow and comprehensiveness. This study employed purposive sampling to select nursing professionals who had participated in TCM specialist nurse training activities. The interviewees were composed exclusively of specialists from six provincial TCM training bases, ensuring both representativeness and feasibility. The complete interview guide is available in Table S2.1 of Supplementary File 2. A total of 20 TCM nursing professionals were interviewed for this study, comprising 7 nursing management experts (35%), 2 nursing education experts (10%), and 11 clinical nursing experts (55%). The basic information of the interview participants is presented in Table S2.2 of Supplementary File 2. Two trained researchers conducted the interviews, either online or in person. Participation was voluntary and based on written

informed consent. Interviews continued until thematic saturation was achieved, ie, when no new indicators or dimensions emerged.[18] Interview transcripts were independently coded by two researchers using thematic analysis. Newly identified indicators were compared against the literature-derived list. Indicators with conceptual overlap were merged, those requiring clearer wording were revised, and newly emerging indicators were retained following team discussion. The refined indicators were then integrated with literature-derived indicators to form the initial indicator pool for the Delphi survey. The detailed results were presented in Table S1.2 of Supplementary File 1.

Expert Panel Selection

To maintain independence across phases, the experts participating in the Delphi and AHP stages were distinct from the interview participants. Experts were recruited using purposive sampling through literature screening, recommendations from the Guangdong Nursing Association, and peer nomination by senior specialists. The rigor of a Delphi study hinges on the scientific selection of experts.[19] Delphi studies for indicator development typically prioritize the depth and relevance of expertise over large sample sizes, and panels of approximately 10–30 experts are commonly considered sufficient to achieve stable consensus in this type of research.[20] On this basis, 16 experts were invited to ensure both diversity of perspectives and sufficient depth of expert review. The panel covered Chinese medicine nursing, nursing education, clinical nursing care, and nursing administration, and included experts with at least 10 years of working or research experience in the respective fields.

The inclusion criteria were as follows:[2,5,17](1) Engaged in TCM nursing, education, healthcare, or administration; (2) Possesses a high level of academic expertise, practical experience, or influence in the field; (3) A senior professional title; (4) ≥10 years of work experience (this requirement was relaxed to ≥5 years for individuals holding a postgraduate degree or higher). The purposive nature of sampling was intended to maximize content expertise rather than statistical generalizability. The exclusion criteria were: (1) Failure to respond to the consultation questionnaire within 4 weeks (28 days); (2) Self-rating as "very unfamiliar" with the research content in their response. The purposive nature of sampling was intended to maximize content expertise rather than statistical generalizability.

Data Collection

A two-round Delphi survey was conducted from June to August 2024. Questionnaires were distributed and collected electronically via Email and WeChat. The initial questionnaire was designed based on previous studies.[17,18] The round 1 questionnaire included four sections: study information, indicator ratings, expert characteristics, and expert self-assessment. Experts rated each indicator on a 5-point Likert scale for importance and could provide open-ended comments regarding deletion, addition, merging, reclassification, or wording revision. After round 1, the research team reviewed the quantitative results together with the qualitative comments. Indicators with a mean importance score >3.5 and a coefficient of variation(CV) [(SD/average importance score) × 100%] <0.25 were provisionally retained,[21,22] whereas indicators failing to meet these criteria or receiving repeated critical comments were re-examined. Decisions to retain, revise, merge, add, reclassify, or delete indicators were made through group discussion, with particular attention to conceptual clarity, consistency with the CIPP-Kirkpatrick framework, and the practical characteristics of TCM specialist nurse training.[21] The revised system was then used for round 2 of consultation. In round 2, experts received the revised questionnaire together with a summary of the round 1 group ratings and major comments, enabling them to reconsider their judgments. Indicators were considered to have reached consensus when at least 70% of experts rated them within the agreement range.[23] The Delphi process was considered complete when expert opinions became stable, and no major structural changes were further proposed. The full questionnaires for both rounds are available in Supplementary Figures S1 and S2.

Statistical Analysis

Data analysis was conducted using Microsoft Excel 2019 and IBM SPSS Statistics 25.0, and the web-based SPSSAU platform.[21] Descriptive statistics were used to summarize expert characteristics and item ratings. In this consensus-based indicator development study, methodological quality was evaluated through multiple complementary aspects rather than conventional statistical power analysis. First, literature review and interviews were used to support initial content generation. Second, the Delphi consultation was used to establish content relevance and expert consensus. Third, the

authority coefficient (Cr), CV, and Kendall's coefficient of concordance (W) were used to evaluate the credibility and coordination of expert judgments.[24] The Cr is calculated as Cr = (Cx + Ca + Cs)/3, where Cx represents the basis of judgment, Ca represents the expert's familiarity with the content, and Cs represents the academic influence, and Cr ≥ 0.70 indicates reliable authority.[23,24] Finally, the consistency ratio (CR) in AHP was used to assess the internal consistency of the weighting procedure.

The criteria for indicator retention were a mean importance score >3.5 and a coefficient of variation <0.25,[21,22] supplemented by expert comments and research team discussion.[21] Response rates above 70% were interpreted as acceptable expert engagement.[23,24] A Cr value ≥0.70 indicated acceptable expert authority,[23,24] and a statistically significant Kendall's W indicated non-random agreement among experts.[24] The AHP was applied to determine the relative weights of indicators. Pairwise comparisons were based on the mean importance ratings from the second Delphi round, following the Saaty scale.[17,24,25] Weight coefficients were calculated using the SPSSAU.[22] The CR values below 0.1 were considered acceptable These analyses were intended to support indicator refinement, consensus assessment, and weighting consistency, but not to establish full psychometric validation of the index system. Representative judgment matrices and the complete consistency testing results for all indicators are detailed in Table S3.1, S3.2, and S3.3 of Supplementary File 3.

Pilot Field Application

A pilot field study was conducted at one provincial Chinese medicine training base (Base F) to examine the feasibility of applying the index system in a real training setting. The pilot included 24 trainees, 3 managers, and 8 clinical mentors who completed the September 2024 training session. Quantitative data were retrieved from the base's management system. Rating-based data were collected from trainees, managers, and mentors using 5-point Likert scales. Evaluation occurred at the end of training and at six-month follow-up. All scores were normalized using the min-max method and combined with the indicator weights to generate domain and overall scores. This pilot application was intended to examine feasibility and practical usability rather than to provide full confirmation or long-term validation of the index system.

Results

Basic Information of the Expert

A total of 16 experts from 6 provinces and municipalities in China were invited, with 15 completing both rounds of consultation. The panel comprised senior professionals from leading provincial TCM hospitals and universities, including hospital vice presidents and directors of nursing from Grade-A Tertiary TCM hospitals responsible for TCM nursing quality control, and heads of specialist nurse training programs. Basic information about the expert is presented in Table 2.

Consensus of Consultation

The effective response rates for the two rounds were 93.75% and 100.00%, respectively, indicating strong enthusiasm among the experts. The Cr was 0.965, exceeding the threshold of 0.7. The Kendall's W coefficient increased from 0.275 in round 1 to 0.292 in round 2 (both $P < 0.001$). Concurrently, the range of CV for all indicators narrowed across rounds (see Table 3). These results indicate that acceptable expert authority and progressive convergence of expert opinion.

Refinement and Finalization of the Index System

Based on a synthesis of literature findings and semi-structured interviews, a preliminary evaluation index system consisted of 4 first-level indicators, 12 second-level indicators, and 57 third-level indicators (see Table S1.2 of Supplementary File 1 for the complete list). Following round 1, 1 second-level indicator was added, and 2 were modified; at the third-level, 8 indicators were added, 19 were revised, and 2 were deleted. A notable structural change was the relocation of "Learning Level Assessment" from the product evaluation domain to the process evaluation domain, reflecting expert consensus that knowledge acquisition in Chinese medicine specialist nurse training should be understood as part of the training process rather than as a terminal outcome. The detailed modification process, including specific deletions, modifications, and the rationale for unadopted expert suggestions, is fully documented in Table S4.1, S4.2, and S4.3 of Supplementary Files 4. In

Table 2 Characteristics of Experts Participating in the Delphi Consultation (n = 15)

Characteristic	Category	n (%)
Age (years)	30–40	1 (6.67)
	41–50	8 (53.33)
	>50	6 (40.00)
Years of experience in the current field	10–20	4 (26.67)
	20–30	6 (40.00)
	>30	5 (33.33)
Professional title	Associate senior	2 (13.33)
	Senior	13 (86.67)
Education background	Bachelor's degree	11 (73.33)
	Master's degree	3 (20.00)
	Doctorate	1 (6.67)
Current field of work	Nursing education	3 (20.00)
	Nursing practice	3 (20.00)
	Nursing management	9 (60.00)

Table 3 Indicators of Expert Consensus Across the Two Delphi Rounds

Round	CV Range	Kendall's W	X^2	Df	*P* value
Round 1	0.000~0.313	0.275	296.594	72	<0.001
Round 2	0.000~0.224	0.292	363.636	83	<0.001

Abbreviations: CV, coefficient of variation; df, Degrees of freedom.

round 2, 3 third-level indicators were further refined for clarity and measurability, and the ranking order of several indicators was adjusted. The indicator refinement flowchart is shown in Figure 4. The detailed modification process is provided in Tables S5.1, S5.2, and S5.3 of Supplementary Files 5. This two-round refinement process resulted in a final index system containing 4 first-level, 13 second-level, and 71 third-level indicators. The first and second-level indicators are shown in Table 4, some of the third-level indicators are presented in Table 5, and the detailed third-level indicators can be found in Supplementary Table S2.

Distinctive Features of the Final 71-Indicator System

The final indicator system was not limited to general training outcomes; rather, it captured multiple dimensions of training context, input, process, and product within the integrated CIPP-Kirkpatrick framework. Importantly, the final structure gave explicit prominence to Chinese medicine-specific competencies, including proficiency in appropriate Chinese medicine techniques, clinical reasoning, and execution of specialized Chinese medicine nursing plans. These features indicate that the system extends beyond conventional end-of-training assessment by incorporating both full-process evaluation and discipline-specific competency assessment.

Indicator Weights

All AHP judgment matrices showed acceptable consistency (CR = 0.000~0.053). At the first level, product evaluation had the highest weight (38.73%), followed by process evaluation (27.48%). At the third level, highly weighted indicators were concentrated in core Chinese medicine competencies, including proficiency in applying appropriate Chinese medicine techniques (4.62%), competency in executing specialized Chinese medicine nursing plans (4.18%), and competency in Chinese medicine clinical reasoning (3.01%). These results suggest that the final index system placed particular emphasis on competencies that are central to Chinese medicine specialist nurse training.

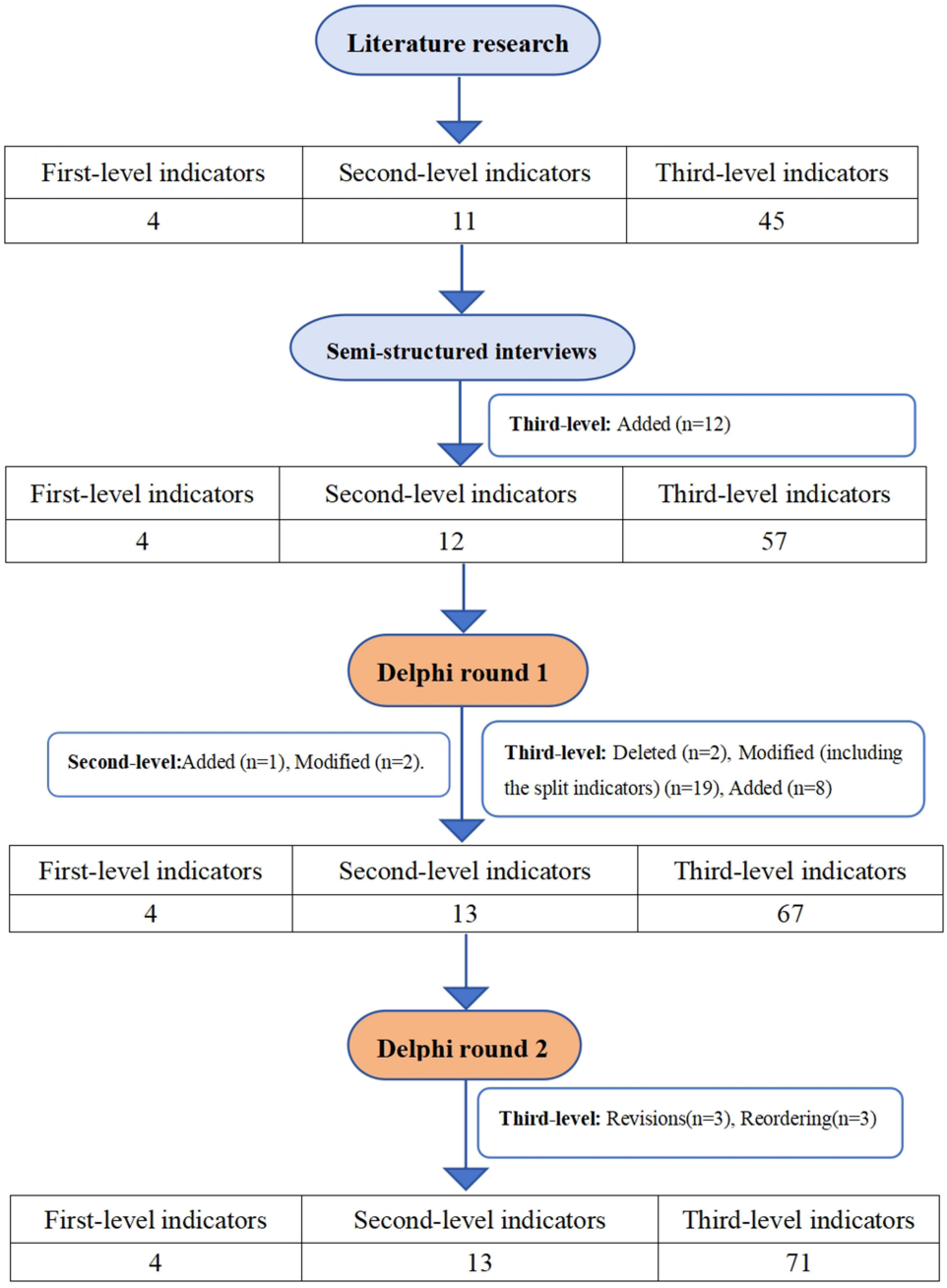

Figure 4 Indicator refinement flowchart.

Pilot Field Application Findings

In the pilot field application at Base F, the total weighted score was 82.55. Among the four primary dimensions, process evaluation showed the highest score (90.99), whereas product evaluation showed the lowest (75.93), with context and input evaluation scoring 85.40 and 78.60, respectively. This pattern suggests that the training base performed relatively well in organizing and implementing the training process, but that longer-term product-oriented outcomes remained comparatively weak. In particular, research-related output and career advancement indicators scored poorly, indicating that the index system may be useful for identifying areas requiring further improvement. These findings should be interpreted as preliminary pilot evidence of feasibility rather than as confirmatory evidence of broad validity.

https://doi.org/10.2147/JMDH.S601386

Table 4 Results of the Second-Round Consultation for First- and Second-Level Indicators

Indicators	Importance Score, Mean ± SD	CV	Weight	Agreement Rate (%)
A. Context evaluation	4.87±0.35	0.072	19.8	100.00
A1 Training needs	4.80±0.56	0.117	3.9	93.33
A2 Training objectives	4.87±0.52	0.106	6.2	93.33
A3 Industry and policy background	5.00±0.00	0.000	9.7	100.00
B. Input evaluation	4.80±0.41	0.086	14.0	100.00
B1 Curriculum construction	4.93±0.26	0.052	3.1	100.00
B2 Teaching conditions	4.93±0.26	0.052	3.1	100.00
B3 Faculty team	4.93±0.26	0.052	3.1	100.00
B4 Rules and regulations	4.93±0.26	0.052	3.1	100.00
B5 Management mechanism	4.87±0.35	0.072	1.6	100.00
C. Process evaluation	4.93±0.26	0.052	27.5	100.00
C1 Training content	4.87±0.52	0.106	6.9	93.33
C2 Teaching evaluation	4.87±0.52	0.106	6.9	93.33
C3 Learning level evaluation	4.93±0.26	0.052	13.7	100.00
D. Product evaluation	5.00±0.00	0.000	38.7	100.00
D1 Behavioral level evaluation	4.87±0.35	0.072	12.9	100.00
D2 Results level evaluation	5.00±0.00	0.000	25.8	100.00

Abbreviation: SD, standard deviation.

Table 5 Consultation Results of Some of the Third-Level Indicators of the Evaluation Index System (Second Round)

Third-level Indicators	Importance Score, Mean ± SD	CV	Weight	Agreement Rate (%)
A1.2 Government policy support	4.80±0.41	0.086	3.24	100.00
A2.2 Training needs for specialized skills operation in TCM nursing	4.93±0.26	0.052	1.12	100.00
A3.3 Training objectives for specialized fields	4.87±0.52	0.106	2.47	93.33
B2.4 Teaching management level	4.80±0.56	0.117	1.61	93.33
B4.4 Operating and management norms of practice bases	4.87±0.52	0.106	1.13	93.33
C1.2 Evaluation of the theory knowledge course for TCM nursing specialization	5.00±0.00	0.000	2.04	100.00
C1.3 Evaluation of the nursing practice skills course for TCM nursing specialization	5.00±0.00	0.000	2.04	100.00
C1.5 Evaluation of the research course	4.80±0.56	0.117	1.08	93.33
C2.2 Feedback evaluation from TCM specialist nurse	4.93±0.26	0.052	2.38	100.00
C2.4 Classroom participation	4.80±0.56	0.117	1.12	93.33
C3.2 Theory examination scores	5.00±0.00	0.000	2.97	100.00
C3.3 Examination scores for TCM appropriate techniques	5.00±0.00	0.000	4.62	100.00
C3.4 Completion number of group Plans	4.87±0.52	0.106	2.18	93.33
C3.5Number of nursing cases completed	4.87±0.52	0.106	2.18	93.33
D1.2 Competency in syndrome differentiation nursing	4.87±0.52	0.106	1.75	93.33
D1.5 Clinical critical thinking ability in TCM	5.00±0.00	0.000	3.01	100.00
D2.1 Patient satisfaction	4.80±0.56	0.117	2.43	93.33
D2.2 Clinical implementation cases of TCM nursing techniques	5.00±0.00	0.000	4.18	100.00
D2.3 Outpatient visits in TCM nursing clinics	5.00±0.00	0.000	4.18	100.00
D2.4 Number of outgoing lecturers	4.93±0.26	0.052	3.09	100.00
D2.5 Number of completed cases of TCM nursing plans for specialized diseases	4.93±0.26	0.052	3.09	100.00
D2.6 Proportion of published papers	4.47±0.64	0.143	1.20	93.33
D2.11 Self-satisfaction	4.80±0.56	0.117	2.43	93.33

Discussion

In this study, we recruited 15 experts from six provinces, which were selected to reflect variation in the timing of initiation of TCM specialist nurse training programs. Among them, 86.7% held senior professional titles, and all possessed over 10 years of experience. The high response rates (93.75% and 100%) and authority coefficient (0.965) support the credibility of the expert consultation process and the quality of expert judgment. Statistical analysis showed that Kendall's coefficient of concordance increased from 0.275 to 0.292 ($P < 0.001$) in the second round, indicating that expert opinions became more consistent. However, these findings should be interpreted as evidence of consensus quality and structural coherence rather than as full validation of the index system.

A key contribution of this study lies in the way the integrated framework was operationalized and contextually adapted for TCM specialist nurse training. Previous evaluation approaches in nursing and health professions education have often emphasized either training outcomes or participant satisfaction, while paying less attention to the relationship between contextual conditions, educational inputs, implementation processes, and downstream competency development. By integrating the CIPP and Kirkpatrick models[26,27] with systematic literature reviews and semi-structured interviews, we ensured the framework covers both educational supply and learner demand, providing a scientifically rigorous foundation for evaluating TCM specialist nurse training. The framework balances process monitoring and outcome orientation. Together, process (27.484%) and outcome (38.734%) evaluations account for over 60% of the total weight. This design aligns with international "Outcome-Based Education" (OBE) requirements while ensuring a closed-loop mechanism.[28]

Regarding specialty characteristics, the system highlights the "mentorship and practice" tradition of TCM nursing. Specifically, "appropriate TCM technology assessment" (D1.4) held the highest composite weight (4.62%), matching evidence-based findings on symptom management[29,30] and reflecting national standardization policies.[31] Other key indicators, such as "outpatient visits" (D2.3) and "syndrome differentiation" (D1.2). These weights encourage a shift from "executing orders" to "active dialectical intervention", aligning with the core competency model proposed by Li et al[32] This ensures that TCM techniques serve clinical practice with precision. This pattern suggests that the index system does more than replicate generic educational evaluation criteria; rather, it embeds competencies that are central to TCM specialist nursing practice. From an educational perspective, this is important because internationally, one of the recurring challenges in specialist nurse training evaluation is that general frameworks may insufficiently capture discipline-specific or culturally embedded competencies.

The pilot field application further illustrated the potential practical value of the framework, but also highlighted the need for cautious interpretation. The relatively high process score suggests that the training base was strong in organizing and delivering the training program, whereas the lower product score indicates that longer-term outcomes may be harder to achieve. In particular, the weak scores for research output, grant applications, and professional promotion suggest that current training may be more effective in transferring immediate clinical skills than in supporting sustained academic development and career transition. This pattern is analytically important because it implies that high-quality training processes do not automatically translate into stronger downstream professional development. For managers and educators, the framework may therefore be useful not only for documenting training performance but also for identifying mismatches between short-term skill acquisition and longer-term developmental outcomes.

These findings suggest that while current training successfully transfers technical skills, it lacks long-term support for career transition. Using this integrated model, managers can clearly identify weaknesses. Future training should adopt a "clinical research mentorship" to help nurses bridge the gap between practice and academic development, establishing robust mechanisms for professional growth. These findings also have broader implications for nursing education and policy. Although the framework was developed for Chinese medicine specialist nurse training in China, the underlying challenge it addresses is not unique to this setting. Many specialist nurse and health professions training programs internationally continue to struggle with how to integrate process monitoring, competency assessment, and longer-term outcome evaluation into a single coherent framework. While evaluation frameworks such as the Kirkpatrick Model and competency-based approaches are widely used internationally, existing studies have largely focused on general training outcomes within standardized clinical contexts. In contrast, our findings highlight the importance of incorporating context and process-sensitive indicators, which are often underrepresented in the global literature. This suggests that

internationally established evaluation models may require contextual adaptation rather than direct transfer, particularly in discipline-specific and culturally embedded practice settings.

This study has several methodological and practical strengths that support its contribution to the field. First, the CIPP-Kirkpatrick integration is theoretically grounded and operationally novel for specialist nurse training contexts. Second, the multi-phase, mixed-methods design enhanced the content validity and practical relevance of the indicators. The high expert authority coefficient (Cr = 0.965) and progressive convergence of expert consensus (Kendall's W increasing from 0.275 to 0.292) further strengthen confidence in the rigor of the consensus process. Third, the pilot application demonstrated practical feasibility and diagnostic utility, identifying specific gaps between process quality and long-term outcomes. Fourth, the framework explicitly embeds Chinese medicine-specific competencies, distinguishing it from generic evaluation tools and enhancing international transferability to other discipline-based training programs.

Limitations

Despite its methodological rigor, this study has several limitations. First, although the index system was informed by literature review, interviews, expert consultation, and pilot field application, it has not yet undergone full psychometric validation or multicenter testing. Therefore, the current evidence supports its content relevance, consensus-based refinement, weighting consistency, and preliminary feasibility rather than comprehensive measurement validation. Its empirical validity and reliability require further testing through large-scale, real-world applications. Second, the field application was conducted at only one training base, which limits the strength of conclusions regarding broader applicability across institutional contexts. Third, the interview participants and Delphi experts were recruited purposively, and the regional concentration of participants may have introduced selection bias and limited external validity. Fourth, the follow-up period in the pilot application was relatively short, which restricted the assessment of longer-term outcomes such as research development and career progression.

Future research should expand expert participation to ensure broader regional diversity and professional perspectives. Subsequent studies should focus on refining quantitative scoring standards and performing multi-center pilot testing. These steps are necessary to fully operationalize the system and observe its long-term impact on the professional development of TCM specialist nurses. In addition, cross-regional and cross-country adaptation studies may help determine whether the framework can be transferred, with contextual modification, to other specialist nurse education systems.

Conclusion

This study developed a theoretical framework that synthesizes the CIPP and Kirkpatrick models to evaluate TCM specialist nurse training. The study's main contribution lies in translating a multidimensional educational evaluation model into a context-sensitive set of indicators that captures both training processes and core Chinese medicine nursing competencies. The pilot field application provided preliminary evidence of feasibility, but the system should still be regarded as an early-stage evaluative framework rather than a fully validated tool. Further multicenter validation, broader real-world application, and cross-context testing are needed before wider implementation. Ultimately, this framework supports the advancement of standardized, evidence-based, and quality-aligned training for TCM specialist nurses.

Data Sharing Statement

The datasets generated and/or analysed during the current study are not publicly available due to the privacy of the expert participants. Still, the data that support the findings of this study are available from Professor Yinqin Zhong upon reasonable request.

Ethical Approval

This study was conducted in accordance with the Declaration of Helsinki. The protocol was reviewed by the Medical Ethics Committee of Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine, which determined that formal ethics approval/full board review was waived because the study involved educational quality improvement activities and the use of de-identified data rather than clinical intervention or collection of biological samples. The committee determined that the study did not involve alteration of patient care, experimental treatment, or other

procedures requiring full board ethics review. Despite the waiver, all participants provided written informed consent for voluntary participation and for the use of their interview responses or evaluation data in the study.

Acknowledgments

The authors would like to thank Professor Hongxia Liu from Beijing University of Chinese Medicine for her valuable guidance and support during the development of this study. We also acknowledge Shenzhen Hospital (Futian) of Guangzhou University of Chinese Medicine for providing the research setting and administrative support for study implementation. All expert consultation, interviews, and evaluation activities were conducted voluntarily and without administrative coercion, employment-related pressure, or performance-related requirements.

Funding

This study was supported by Futian District Health Bureau Research project [grant number FTWS003], as well as the "2024 Annual Higher Education Scientific Research Planning Project" of the China Association of Higher Education [Approval Number 24JY0402]. The funding sources had no role in the study design, data collection, data analysis, interpretation of the findings, manuscript preparation, or the decision to submit the manuscript for publication.

Disclosure

The authors declare that they have no financial or non-financial competing interests related to this study.

References

1. Litchman ML, Schlepko T, Rowley T, McFarland M, Fiander M. A scoping review of advanced practice registered nurses consensus model outcomes: part four of a four-part series on critical topics identified by the 2015 nurse practitioner research agenda. *J Am Assoc Nurse Pract*. 2018;30(12):710–723. doi:10.1097/jxx.0000000000000111
2. Gilligan L, Page E, Hall J, et al. Diabetes specialist nurse support, training and 'virtual' advice reduces district nurse visits and improves outcomes for people with diabetes requiring visits for insulin administration. *Diabetes Res Clin Pract*. 2025;219:111948. doi:10.1016/j.diabres.2024.111948
3. Sili A, Zaghini F, Monaco D, et al. Specialized nurse-led care of chronic wounds during hospitalization and after discharge: a randomized controlled trial. *Adv Skin Wound Care*. 2023;36(1):24–29. doi:10.1097/01.ASW.0000897444.78712.fb
4. Zhang Q, Sharan A, Espinosa SA, Gallego-Perez D, Weeks J. The path toward integration of traditional and complementary medicine into health systems globally: the World Health Organization report on the implementation of the 2014–2023 strategy. *J Altern Complement Med*. 2019;25 (9):869–871. doi:10.1089/acm.2019.29077.jjw
5. Zhang S, Gai H, Zhang H, et al. Analysis and prospect of the current situation, training, and use of traditional Chinese medicine specialty nurses. *Nurs Pract Res*. 2024;24(9):1281–1285. doi:10.3969/j.issn.1672-1756.2024.09.001
6. Li Y, Li C, Shi JH, et al. Construction of a core competency evaluation system for traditional Chinese medicine nurses. *J New Chin Med*. 2023;55 (21):182–187.
7. Si HJ. Research on the Current situation and Influencing Factors of the Core Competence of Traditional Chinese Medicine Nurses in Hunan Province. Hunan University of Chinese Medicine; 2024.
8. Zhang YY, Li Y, Wang XQ, Zhao RQ, Su J, Duan PB. Construction of a progressive training program for specialized traditional Chinese medicine nursing nurses in Jiangsu Province. *Chin Nurs Manag*. 2024;24(12):1824–1830.
9. Long L, Wei L, Wang Q, et al. Visual analysis of hotspot and trend of nurses specializing in traditional Chinese medicine in China based on CiteSpace software. *Gen Nurs*. 2023;21(17):2305–2309. doi:10.12104/j.issn.1674-4748.2023.17.001
10. Singh MD. Evaluation framework for nursing education programs: application of the CIPP model. *Int J Nurs Educ Scholarsh*. 2004;1:Article13. doi:10.2202/1548-923x.1023
11. Ullah H, Huma S, Yasin G, et al. Curriculum and program evaluation in medical education: a short systematic literature review. *Ann Med Surg*. 2024;86(10):5988–5994. doi:10.1097/ms9.0000000000002518
12. Dubrowski A, Morin MP. Evaluating pain education programs: an integrated approach. *Pain Res Manag*. 2011;16(6):407–410. doi:10.1155/2011/320617
13. McClintock AH, Starks H, Williams M. Women's health for a primary care workforce. *Clin Teach*. 2022;19(3):251–256. doi:10.1111/tct.13483
14. Wang C. *Research on the construction of a standardized training effect evaluation index system for newly hired nurses based on the CIPP model and Kirkpatrick model* [Master's thesis]. Nanjing University; 2019.
15. Jiang N, Liu WB, Zong Y, Yu L, Cheng S. Construction of an index system of core competence assessment for sleep medicine nurse specialists in China: a Delphi study. *BMC Nurs*. 2024;23(1):689. doi:10.1186/s12912-024-02349-2
16. Jing CY, Liu RX, Chu HL, Sun WW, Liao X. Interpretation of standards for conducting and reporting Delphi studies (CREDES) in healthcare research. *Chin J Evid Based Med*. 2023;23(2):233–239. doi:10.7507/1672-2531.202210047
17. Wang J. *Construction and preliminary application of the evaluation index system for core competitiveness in nursing undergraduate talent training based on the CIPP model* [Master's thesis]. Nanchang University; 2022.
18. Mei Y. Research on the Concept Framework and Assessment Tools of Benefit Sensation of Stroke Caregivers. Zhengzhou University; 2018.
19. McPherson S, Reese C, Wendler MC. Methodology update: delphi studies. *Nurs Res*. 2018;67(5):404–410. doi:10.1097/nnr.0000000000000297
20. Liang CH, Yin GX, Wang YQ, et al. Methodological processes and recommendations for the application of the delphi method in the development and revision of clinical practice guidelines for traditional chinese medicine. *Med J PUMCH*. 2023;14(5):1084–1090.

21. Ma H. *Construction and implementation of clinical ladder model on competency-based continuing professional education among military nurses* [Master's thesis]. Army Medical University of the PLA; 2022.
22. Zhou J, Ma S. *SPSSAU Research Data Analysis Methods and Applications*. 1st ed. Electronics Industry Press; 2024.
23. Chen C, Zheng X, Zhang M, et al. Developing a crisis leadership evaluation system for Chinese nursing staff during major infectious disease emergencies: a modified delphi study. *BMC Nurs*. 2025;24(1):423. doi:10.1186/s12912-025-03050-8
24. Shen L, Yang J, Jin X, Hou L, Shang S, Zhang Y. Based on delphi method and analytic hierarchy process to construct the evaluation index system of nursing simulation teaching quality. *Nurse Educ Today*. 2019;79:67–73. doi:10.1016/j.nedt.2018.09.021
25. Lucas RI, Promentilla MA, Ubando A, Tan RG, Aviso K, Yu KD. An AHP-based evaluation method for teacher training workshop on information and communication technology. *Eval Program Plann*. 2017;63:93–100. doi:10.1016/j.evalprogplan.2017.04.002
26. Huang X, Wang R, Chen J, et al. Kirkpatrick's evaluation of the effect of a nursing innovation team training for clinical nurses. *J Nurs Manag*. 2022;30(7):2165–2175. doi:10.1111/jonm.13504
27. Toosi M, Modarres M, Amini M, Geranmayeh M. Context, input, process, and product evaluation model in medical education: a systematic review. *J Educ Health Promot*. 2021;10(1):199. doi:10.4103/jehp.jehp_1115_20
28. Harden RM. Outcome-based education: the future is today. *Med Teacher*. 2007;29(7):625–629. doi:10.1080/01421590701729930
29. Zhang L, Qin YM, Zheng LX, et al. Governor vessel moxibustion: ancient Chinese medical technology with new vitality. *Chin J Integr Medi*. 2017;23(5):396–400. doi:10.1007/s11655-015-2149-7
30. Zhu ZG, Wang JR, Pan XY. Efficacy of scraping therapy on blood pressure and sleep quality in stage I and II essential hypertension: a systematic review and meta-analysis. *J Integr Med*. 2024;22(1):12–21. doi:10.1016/j.joim.2023.11.006
31. Notice on the issuance of the action plan for further improving nursing services (2023–2025) (Bulletin of the National Health Commission of the People's Republic of China) 5-8 (2023).
32. Li L. Development and Implementation of a Training Program for Traditional Chinese Medicine Oncology Nurse Specialists. Shaanxi University of Chinese Medicine; 2018.