

Assessment Tools for Innovation Ability Among Nursing Students: A Systematic Review

Zhe-Hao Hu

Beijing University of Chinese Medicine

Xin-Yu Zhao

Shandong University

Si-Qi Ye

Shandong University

Xin-Yue Zhang

Beijing University of Chinese Medicine

Mei-Qi Meng

Beijing University of Chinese Medicine

Xue-Jing Li

Beijing University of Chinese Medicine

Yu-Fang Hao

bucmnursing@163.com

Beijing University of Chinese Medicine

Research Article

Keywords: nursing students, innovation ability, assessment tools, systematic review, reliability and validity

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10802194/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Innovation is an essential competency for nurses in increasingly complex healthcare environments. Accurate assessment of nursing students' innovation ability is crucial for identifying learning needs, evaluating educational interventions, and informing curriculum development. However, existing assessment tools are heterogeneous, often lack nursing specificity, and provide insufficient evidence of structural validity. The evolutionary relationships among these tools also remain unclear.

Aims

To systematically review assessment tools for nursing students' innovation ability, evaluate their psychometric properties, applicability, and limitations, and provide evidence-based recommendations for tool selection and future development.

Methods

PubMed, CINAHL, Embase, CNKI, and Wanfang Data were searched from inception to March 2026. Studies involving nursing students and reporting the development, revision, or application of innovation ability assessment tools were included. Data were extracted using a standardized form, and tools were evaluated in terms of reliability, validity, applicability, and practicality.

Results

Ten assessment tools were identified and classified into four types: creativity/personality tendency tests ($n = 1$), comprehensive innovation ability questionnaires ($n = 3$), innovative behavior scales ($n = 4$), and research/innovation and entrepreneurship scales ($n = 2$). All tools reported internal consistency, with 90% demonstrating good reliability (Cronbach's $\alpha \geq 0.80$). However, only four tools provided relatively complete structural validity evidence, while 60% reported neither exploratory nor confirmatory factor analysis. Innovative behavior scales showed a clear evolution from Scott and Bruce's 1994 unidimensional scale to multidimensional and context-specific adaptations. The Chinese version of the Creativity and Innovation Effectiveness Profile (CIEP) and the Nurse Innovative Behavior Scale (NIBS) demonstrated the strongest nursing specificity.

Conclusion

Existing tools generally demonstrate satisfactory reliability but insufficient validity evidence, particularly structural validity. Future research should prioritize rigorous structural validation, nursing-specific instrument development, and stage-specific, multi-level assessment systems tailored to students' developmental needs.

1. Introduction

As shown in Fig. 1., In today's rapidly changing and increasingly complex healthcare environments, nurses are frequently required to respond to unexpected situations across diverse settings. As a critical element of the healthcare system, they must provide holistic care while adapting to evolving patient needs[1]. Innovation is no longer an optional attribute for nurses — it is a professional imperative. Contemporary healthcare systems demand that nurses move beyond the faithful execution of established protocols to become active contributors to practice improvement, whether by redesigning care delivery workflows, creating novel patient education tools, or identifying inefficiencies in clinical processes that have gone unchallenged for decades. The capacity to innovate, once considered the domain of a gifted minority, is now recognized as a competency that can — and must — be systematically cultivated through nursing education[2].

Nursing students, as the future workforce, stand at the center of this imperative. Yet cultivating innovation ability depends on a prerequisite that is often overlooked: the capacity to assess it accurately. Without valid and reliable assessment tools, educators cannot identify students' strengths and deficits, evaluate the effectiveness of educational interventions, or make evidence-informed decisions about curriculum design. The quality of assessment directly determines the quality of education. In educational and health measurement, the quality of assessment directly determines the validity of inferences drawn from performance data and, ultimately, the quality of education itself [3, 4].

It is at this foundational level that the current literature reveals a troubling gap. A growing body of empirical research has examined nursing students' innovation ability, investigating its correlates with self-efficacy [5], professional commitment [6], psychological capital [7], and other educational variables. Yet far fewer studies have rigorously examined the measurement properties of the instruments that generate these data. The field appears to be building a rich body of correlational knowledge on a psychometric foundation that has not been systematically inspected. Existing reviews in this area remain limited in scope: Zhang et al. (2024) [8], for example, included only seven instruments in their narrative review and did not trace the evolutionary relationships among tools or critically appraise structural validity evidence using standardized psychometric frameworks.

Three specific methodological challenges characterize the current landscape. First, instrument selection lacks domain specificity. Many studies uncritically adopt general creativity or innovation scales developed in psychology or management disciplines. For instance, the Torrance Tests of Creative Thinking (TTCT), developed in 1960s America and requiring complex, culturally embedded scoring procedures, continues to be applied to Chinese nursing students

despite its questionable content and contextual validity in clinical nursing environments [9, 10]. Second, even well-validated generic innovation scales are disconnected from the clinical realities of nursing practice. Nursing innovation often manifests in domain-specific forms — such as optimizing bed-side workflows, adapting patient safety protocols, or designing assistive devices — that generic instruments were never designed to capture [8]. Third, and most fundamentally, the psychometric evidence supporting these tools is highly variable and often incomplete. While internal consistency (e.g., Cronbach's alpha) is routinely reported, structural validity evidence — which addresses the essential question of whether an instrument accurately reflects the theoretical dimensionality of the construct being measured — is frequently absent or insufficiently evaluated [11, 12]. No existing review has systematically appraised this evidence base while simultaneously mapping the evolutionary trajectories through which generic instruments have been progressively adapted for nursing contexts.

To address these gaps, this systematic review was conducted with three primary objectives: (1) to comprehensively identify and categorize assessment tools used to evaluate nursing students' innovation ability; (2) to critically appraise their psychometric properties — with particular attention to the under-examined domain of structural validity — across the dimensions of reliability, validity, cross-cultural applicability, and practicality using established psychometric standards; and (3) to trace the evolutionary trajectories of these tools (particularly innovative behavior scales) to provide both practical evidence-based recommendations for tool selection and a conceptual roadmap for future instrument development in nursing education.

2. Methods

2.1. Protocol and Registration

This systematic review was conducted and reported in strict accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [13]. The protocol was prospectively registered on the Open Science Framework (OSF) (DOI: 10.17605/OSF.IO/EHCXM).

2.2. Search Strategy

PubMed, CINAHL, Embase, CNKI, and Wanfang Data were systematically searched from database inception to March 2026. To minimize publication bias and capture comprehensively validated or modified instruments, grey literature (including doctoral dissertations, conference proceedings, and institutional reports) was also searched [14]. (Full Search Strategies are on Supplementary Materials)

The Chinese database search strategy (using CNKI as an example) employed subject term and keyword combinations including terms for "nursing students" (e.g. 护理学, 护理专业, 护理本科), "innovation ability" (e.g. 创新能力, 创新力, 创新意识, 创新思维), and "assessment tools" (e.g. 量表, 问卷, 量表, 量表).

The English database search strategy (using PubMed as an example) combined MeSH terms and title/abstract keywords: ("Students, Nursing"[MeSH] OR "nursing student"[Title/Abstract] OR "nursing students"[Title/Abstract] OR "nursing undergraduate"[Title/Abstract]) AND ("Creativity"[MeSH] OR "innovative behavior"[Title/Abstract] OR "innovation capability"[Title/Abstract] OR "creative thinking"[Title/Abstract] OR "creative ability"[Title/Abstract]) AND ("scale"[Title/Abstract] OR "questionnaire"[Title/Abstract] OR "instrument"[Title/Abstract] OR "assessment tool"[Title/Abstract] OR "measurement tool"[Title/Abstract] OR "Educational Measurement"[MeSH]).

2.3. Inclusion and Exclusion Criteria

Inclusion criteria were: (1) participants were nursing students at any educational level (including diploma, associate, baccalaureate, and postgraduate); (2) content focused on the development, revision, or application of innovation ability assessment tools; and (3) basic characteristics of the assessment tool were clearly reported.

Exclusion criteria were: (1) studies involving non-nursing participants; (2) studies that only constructed an evaluation index system without developing a concrete assessment tool; (3) duplicate publications or review articles; and (4) studies with inaccessible full texts or critical missing information.

2.4. Study Selection and Data Extraction

Two researchers independently screened all retrieved records by title, abstract, and full text. Disagreements were resolved through discussion or consultation with a third senior researcher. The study selection process strictly followed the four-phase PRISMA flow framework (identification, screening, eligibility, and inclusion) [15]. Data were extracted using a standardized form adapted from the COSMIN data extraction template [16], collecting: author, publication year, target population, tool name, origin/theoretical framework, item counts, dimensional structure, scoring format, and detailed psychometric properties (reliability and validity).

2.5. Quality Appraisal and Psychometric Evaluation

The methodological quality and psychometric properties of the included instruments were critically appraised using criteria informed by the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines and established quality criteria for measurement properties [12, 16, 17]:

(1) Reliability: Internal consistency was evaluated via Cronbach's alpha or McDonald's omega ($\alpha/\omega \geq 0.70$ considered acceptable, ≥ 0.80 considered good) [12]; test-retest reliability was evaluated using intra-class correlation coefficients (ICC ≥ 0.70 considered adequate) where available.

(2) Validity: Content validity was assessed based on expert panel evaluation (e.g., Item-level Content Validity Index, I-CVI ≥ 0.78); structural validity was appraised based on reported factor analytical procedures (exploratory factor analysis, EFA, and confirmatory factor analysis, CFA with standard model fit indices such as RMSEA < 0.08 and CFI/TLI > 0.90) [12, 18].

(3) Applicability/Feasibility: Evaluated based on instrument relevance to nursing clinical/educational contexts, language readability, and specific validation in nursing student cohorts [16].

(4) Practicality: Assessed by item count, estimated completion time, and scoring convenience.

Two reviewers independently performed all quality appraisals, with inter-rater agreement calculated and discrepancies resolved via consensus. Quality appraisal results are presented narratively and in structured comparative tables.

2.6. Data Synthesis

- A combination of narrative synthesis and comparative content analysis was performed, guided by the framework for systematic review synthesis of measurement tools [18]. First, included tools were categorized by construct focus (e.g., creative thinking, innovative behavior, innovative literacy). Second, key characteristics of each tool were extracted and tabulated systematically. Third, cross-sectional comparative analysis was conducted to contrast domain specificity, dimension structures, and psychometric completeness between domestic (Chinese) and international tools. Fourth, a longitudinal evolutionary mapping was executed to trace the lineage and developmental trajectory of foundational tools (e.g., generic innovative behavior scales) as they were adapted and modified for nursing education over time.

3. Result

3.1. Study Selection

The initial search yielded 442 records. After removing duplicates, 255 records remained. Following title and abstract screening, 187 records were excluded, and 54 full-text articles were assessed for eligibility. After full-text review, 44 articles were excluded for not meeting the inclusion criteria. Ultimately, 10 studies involving 10 assessment tools were included. The study selection flow diagram is presented in Fig. 2.

3.2 Basic Characteristics of Included Tools

The basic characteristics of the 10 included tools are summarized in Table 1, including tool name, first author, publication year, number of items, dimensions, and so on.

Table 1
Basic characteristics of the included assessment tools

Tool name	First author	Year	Study population	Scoring method	Study region	Sample size	Items/ Dimensions	Dimensions	Completion time	Test-retest interval	Source
CAP	Lin Xingtai [16]	1986	Nursing undergraduates [17]	Likert 3-point	Shandong Province	402	50/4	Risk-taking, Curiosity, Imagination, Challenge	NR	NR	Cross-adapted
ICEA	Jon Warner [7]	2002	Nursing undergraduates [18]	Likert 5-point	Guangdong Province	154	84/7	Innovation awareness, Curiosity, Breaking fixed thinking patterns, Idea cultivation ability, Willingness to experiment/take risks, Courage & psychological resilience, Energy persistence	10 min	NR	Original applied
CIEP	Ye Siqi [8]	2025	Nursing undergraduates [8]	Likert 5-point	Beijing, Hebei Province	531	31/4	Creative awareness, Pattern-breaking skills, Critical thinking ability, Idea cultivation ability	20–30 min	14 days	Adapted
CSIAQ	Wang Qiqi [9]	2012	Nursing undergraduates [19]	Likert 5-point	Shanxi province	246	21/3	Innovative thinking ability, Innovative learning ability, Innovative practice ability	NR	NR	Original applied
NIBS	Bao Ling [10]	2012	Nursing interns [20]	Likert 5-point	Hubei Province	506	10/3	Idea generation, Support acquisition, Idea realization	≥ 180 s	NR	Original applied
CSIBS	Jiang Liping [11]	2018	Vocational nursing students [21]	Likert 5-point	Anhui Province	446	17/4	Thinking style, Academic exploration, Life practice, Academic learning	NR	NR	Original
CSIBS	Xin Tiantian [12]	2016	Nursing undergraduates [22]	Likert 5-point	Hunan Province	338	29/5	Student support, Organizational support, Supervisor support, Innovative self-efficacy, Innovative behavior, Constructive debate	NR	NR	Original applied
IIBS	Zhao Zhigang [13]	2011	Nursing undergraduates [23]	Likert 5-point	Hubei Province	215	9/3	Innovative idea generation, Innovative idea promotion, Innovative idea realization	NR	NR	Original applied
IESS	Shen Yijing [14]	2021	Nursing undergraduates [14]	Likert 5-point	Jiangsu Province	263	17/4	Innovation & entrepreneurship awareness status, Innovation & entrepreneurship cognition status, Innovation & entrepreneurship ability, Innovation & entrepreneurship education & professional identity	NR	NR	Newly developed

Tool name	First author	Year	Study population	Scoring method	Study region	Sample size	Items/ Dimensions	Dimensions	Completion time	Test-retest interval	Source
ISQ	Wang Jing [15]	2014	Postgraduate nursing students [15]	Likert 4-point	Nationwide	164	33/4	Thinking characteristics, Personality characteristics, Behavioral characteristics, Knowledge structure	NR	NR	Newly c

Based on measurement content, the tools were categorized into four types: (1) Classic creativity/personality tendency tests (n = 1): the Chinese version of the Williams Creativity Assessment Packet (CAP)[19]; (2) Comprehensive innovation ability self-report questionnaires (n = 3): the Innovation & Creativity Effectiveness Assessment (ICEA)[20], the Chinese version of the Creativity and Innovation Effectiveness Profile (CIEP)[21], and the College Students' Innovation Ability Questionnaire (by Wang Qiqi)[22]; (3) Innovative behavior scales (n = 4): the Nurse Innovative Behavior Scale (NIBS)[23], the College Students' Innovative Behavior Scale (by Jiang Liping)[24], the Innovative Behavior Scale (by Xin Tiantian)[25], and the Individual Innovative Behavior Scale (by Zhao Zhiqing)[26]; and (4) Research/innovation and entrepreneurship scales (n = 2): the Innovation and Entrepreneurship Status Scale for Nursing Undergraduates (by Shen Yijing)[27] and the Innovation Status Questionnaire for Nursing Postgraduates (by Wang Jing)[28].

3.3 Quality Appraisal Results

The CIEP covers all four evaluation dimensions without any omissions; in contrast, other assessment tools have gaps in their coverage. Quality appraisal results are presented in Table 2.

Table 2
Quality appraisal of the included assessment tools

Tool name	Reliability	Validity	Applicability	Practicality
CAP	0.92 / 0.872	NR / NR	Widely used in nursing, not nursing-specific	50 items / NR / 3-point
ICEA	0.975 / NR	NR / NR	Generic scale, weak nursing specificity	84 items / 10 min / 5-point
CIEP	0.894 / 0.846	0.964 / CFA good	Nursing student-specific, culturally adapted	31 items / 20–30 min / 5-point
CSIAQ	0.92 / NR	NR / EFA acceptable	Validated in nursing students, not nursing-specific	21 items / NR / 5-point
NIBS	0.775–0.898 / NR	NR / NR	Nursing student/nurse-specific	10 items / ≥ 180 s / 5-point
CSIBS (Jiang)	0.961 / NR	NR / NR	Validated in nursing students	17 items / NR / 5-point
CSIBS (Xin)	0.864 / NR	NR / NR	Validated in nursing students	29 items / NR / 5-point
IIBS	0.87 / 0.96	0.96 / NR	Validated in nursing students	9 items / NR / 5-point
IESS	0.781–0.923 / NR	0.96 / CFA + EFA good	Nursing student-specific	17 items / NR / 5-point
ISQ	0.809 / 0.821	0.821 / NR	Postgraduate nursing student-specific	33 items / NR / 4-point

3.3.1 Reliability: Generally Satisfactory

All included tools reported internal consistency reliability (Cronbach's α). Nine tools (90%) achieved $\alpha \geq 0.80$, indicating good reliability. The ICEA demonstrated the highest reliability ($\alpha = 0.975$), while the Wang Jing questionnaire showed relatively lower but still acceptable reliability ($\alpha = 0.809$). However, only four tools reported test-retest reliability: the Williams CAP (0.872), the CIEP (ICC = 0.846), the Individual Innovative Behavior Scale (0.96), and the Wang Jing questionnaire (0.821). The remaining six tools did not report test-retest reliability.

3.3.2 Validity: Generally Insufficient

Only four tools (CIEP, Wang Qiqi's questionnaire, Shen Yijing's scale, Wang Jing's questionnaire) reported relatively complete content or structural validity evidence. The CIEP demonstrated the most rigorous validity testing: its content validity index was 0.964, and confirmatory factor analysis (CFA) showed good model fit (CFI = 0.925, RMSEA = 0.051). Shen Yijing's scale also demonstrated good CFA fit indices (CFI = 0.990, RMSEA = 0.031). Wang Qiqi's questionnaire reported exploratory factor analysis (EFA) results with a cumulative variance contribution rate of 52.601%. However, 60% of the tools (including the ICEA, NIBS, Jiang Liping's scale, and Xin Tiantian's scale) did not report either EFA or CFA, leaving their structural validity unsubstantiated.

3.3.3 Nursing Specificity: Considerable Variation

In terms of nursing specificity, the CIEP and NIBS demonstrated the strongest performance. The CIEP underwent systematic cross-cultural adaptation and validation using a sample of 531 nursing undergraduates from Beijing and Hebei Province, with item content reviewed by nursing education experts. The NIBS, specifically developed for nursing professionals, contains three dimensions closely aligned with clinical nursing practice—"idea generation," "support acquisition," and "idea realization"—and has been widely used among nursing students and clinical nurses. In contrast, generic tools such as the ICEA (84 items, 7 dimensions), despite their excellent reliability, contain items conceptually oriented toward psychology (e.g., "breaking fixed thinking patterns," "courage and psychological resilience") with limited relevance to nursing clinical practice.

3.3.4 Practicality: Marked Differences in Feasibility

In terms of item count, the Individual Innovative Behavior Scale (9 items) and NIBS (10 items) were the most concise and thus suitable for large-scale rapid assessment. Conversely, the ICEA (84 items) imposed a substantial response burden, and the Wang Jing questionnaire (83 items in total) was not conducive to widespread application. Regarding scoring methods, most tools employed a Likert 5-point scale, with the exceptions of the Williams CAP (3-point) and the Wang Jing questionnaire (4-point). More than half of the tools did not report completion time.

3.4 Evolutionary Trajectory of Innovative Behavior Scales

The innovative behavior scales demonstrated a clear evolutionary trajectory (Fig. 3.). The origin can be traced to Scott and Bruce's (1994)[29] 6-item unidimensional Individual Innovative Behavior Scale (IIBS), developed based on social exchange theory to measure innovative behavior in workplace settings. Janssen (2000) [30] extended this into the 9-item, three-dimensional Innovative Work Behavior (IWB) scale, comprising idea generation, idea promotion, and idea realization, using a Likert 7-point scale. Subsequently, Chinese scholars undertook successive nursing-specific adaptations: Zhao Zhiqing et al. (2011) [26] revised it into a 9-item Chinese version suitable for nursing students, converting to a Likert 5-point scale. Bao Ling et al. (2012) [23] developed the 10-item NIBS specifically for nursing professionals. Xin Tiantian (2016) [25] further expanded the scale into a 25-item, five-dimensional version incorporating student support, organizational support, supervisor support, innovative self-efficacy, and innovative behavior, thereby integrating environmental support and self-efficacy dimensions. Jiang Liping (2018) [24] developed a 17-item, four-dimensional version focused on academic contexts, incorporating characteristics of university students' learning and daily life. This evolutionary trajectory reflects a progressive deepening from generic workplace contexts to specific professional contexts and from unidimensional to multidimensional measurement.

4. Discussion

4.1 The Structural Deficit: "Good Reliability, Insufficient Validity"

This systematic review revealed a striking pattern: current assessment tools for nursing students' innovation ability demonstrate generally good reliability but widespread insufficient validity evidence — particularly structural validity. While all tools reported internal consistency and 90% achieved good levels (Cronbach's $\alpha \geq 0.80$), 60% had not undergone rigorous factor analysis. This finding directly aligns with Deng et al. (2025) in their systematic review of clinical nursing teacher competency tools and echoes broader measurement concerns in healthcare education [31]. From a psychometric perspective, high internal consistency is a necessary but insufficient condition for validity; a high Cronbach's alpha can merely reflect item redundancy rather than structural integrity [4, 32]. Good reliability in the absence of validity evidence means we cannot determine whether the dimensional structure of these tools is theoretically sound or whether items genuinely represent the construct of "nursing innovation ability" [16, 33]. For example, while the ICEA demonstrated excellent reliability ($\alpha = 0.975$), the rationale for its 84 items and seven dimensions lacks structural validity support; whether dimensions such as "energy persistence" truly fall within the construct domain of innovation capability remains debatable. Taken together, these findings corroborate the critical issues identified in our introduction: uncritical borrowing of generic instruments and uneven psychometric evidence that undermines the foundation of a growing body of correlational research.

4.2 The Dilemma of Tool Selection and Evidence-Based Recommendations

Our findings suggest that the current dilemma in tool selection stems from two interrelated tensions.

First is the tension between "generic" and "domain-specific" measurement. According to Social Cognitive Theory and health measurement guidelines, behavioral capabilities are highly domain-specific [34, 35]. Generic scales (e.g., ICEA, Wang Qi's questionnaire) perform well in reliability but are disconnected from nursing clinical practice, failing to capture context-bound innovative behaviors such as "identifying bedside safety hazards" or "optimizing clinical procedures" [36]. Conversely, nursing-specific tools (e.g., CIEP, NIBS) contain items better aligned with clinical education contexts. However, some (e.g., Shen Yijing's scale validated in a single institution; Wang Jing's questionnaire evaluated only in its development study) lack cross-sample validation and measurement invariance testing, limiting their generalizability.

Second is the absence of a developmental, stratified assessment system. As posited by Benner's "Novice to Expert" continuum and professional competence frameworks [37, 38], nursing students at different educational stages encounter fundamentally distinct learning environments and clinical realities. Junior students primarily engage with basic sciences; assessment at this stage should emphasize creative personality traits and divergent thinking. Senior students transition into clinical practice, where assessment must focus on contextualized clinical innovation behaviors. Postgraduate students must demonstrate advanced capabilities in both research-driven and practice-based innovation. Yet, no stratified assessment battery tailored to these distinct developmental stages currently exists.

Based on our synthesis, we offer three pragmatic recommendations: (1) When rigorous, multidimensional evaluation of innovation capability is required, the CIEP—which possesses the most robust psychometric validation—should be the primary instrument. (2) For rapid screening or educational evaluation in clinical settings, the NIBS is recommended due to its strong nursing specificity and concise structure. (3) For postgraduate nursing cohorts, while Wang Jing's questionnaire offers domain-specific relevance, it requires formal structural validity testing before widespread implementation.

4.3 A Field-Level Imbalance: Tool Development Versus Correlational Research

A noteworthy pattern emerged: the overwhelming majority of published studies on nursing students' innovation ability are cross-sectional surveys analyzing correlates such as self-efficacy, professional commitment, psychological capital, and artificial intelligence readiness. In contrast, studies dedicated to instrument development, cross-cultural adaptation, or psychometric validation are scarce. While correlational studies provide valuable preliminary data, their heavy reliance on unscrutinized measurement tools creates a "measurement schmeasurement" phenomenon—where complex statistical models are built upon fragile psychometric foundations [39]. First, causal relationships cannot be established from cross-sectional designs [40]. Second, variable selection

often follows empirical trends rather than robust theoretical frameworks. Third, and most critically, researchers routinely assume that borrowed scales are valid without confirming content validity or structural validity in their target samples [16, 39]. We argue that the field urgently needs to pivot from "uncritically applying tools" to "rigorously refining tools," directing research efforts toward cross-sample validation, measurement invariance, and continuous optimization of core instruments like the CIEP and NIBS.

4.4 Limitations

This systematic review has several limitations. First, psychometric reporting was subject to secondary transmission bias, as data were extracted from primary studies that recalculated Cronbach's α for convenience without conducting formal validity re-testing. Second, due to complex lineage chains of adaptation and revision, we included 10 core tools with relatively complete psychometric data, which may exclude minor derivative variations. Third, grey literature was not included, which may introduce publication bias [14]. Fourth, although informed by COSMIN standards, our appraisal utilized a synthesized framework rather than the full COSMIN Risk of Bias checklist [16], and did not explicitly differentiate between "property absent" and "property unreported," which may result in a conservative evaluation of certain instruments.

5. Conclusion

This systematic review shows that current assessment tools for nursing students' innovation ability generally demonstrate satisfactory reliability, but validity evidence—particularly structural validity—is insufficient. The tools are diverse in type but vary considerably in nursing specificity. Nursing students at different grade levels have differentiated assessment needs, yet a stratified assessment system remains absent. Future research should prioritize rigorous structural validity testing and large-sample cross-regional validation of existing core tools (particularly the CIEP and NIBS). On this foundation, efforts should be directed toward developing assessment instruments deeply integrated with nursing clinical practice contexts, constructing stage-specific, multi-level assessment systems, and promoting integrated analyses linking assessment results with educational interventions—ultimately serving the improvement of nursing innovation talent cultivation quality.

Declarations

Acknowledgments

Not applicable.

CRediT authorship contribution statement

Zhehao Hu: Writing - review & editing, Writing - original draft, Software, Visualization, Validation, Methodology, Formal analysis, Data curation, Investigation, Resources. **Xinyu Zhao:** Writing - review & editing, Writing - original draft, Software, Visualization, Validation, Methodology, Formal analysis, Data curation, Investigation, Resources. **Siqi Ye:** Methodology, Investigation. **Xinyue Zhang:** Methodology, Investigation. **Meiqi Meng:** Methodology. **Xuejing Li:** Writing-review & editing, Formal analysis, Validation, Supervision, Conceptualization, Project administration, Resources. **Yufang Hao:** Supervision, Conceptualization, Funding acquisition, Project administration, Resources.

Funding

This work was supported by the National Administration of Traditional Chinese Medicine [grant numbers ZYYCXTD—D—202401].

Declaration of competing interest

The authors report no conflicts of interest.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Clinical trial number

Not applicable.

References

1. Chan ZC. A systematic review of creative thinking/creativity in nursing education. *Nurse Educ Today*. 2013;33(11):1382–7. 10.1016/j.nedt.2012.09.005.
2. Cusson RM, Meehan C, Bourgault A, Kelley T. Educating the next generation of nurses to be innovators and change agents. *J Prof Nurs*. 2020;36(2):13–9. 10.1016/j.profnurs.2019.07.004.

3. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–7. 10.1046/j.1365-2923.2003.01594.x.
4. Streiner D, Norman G, Cairney J. Health measurement scales: a practical guide to their development and use (5th edition). *Aust N Z J Public Health*. 2016;40(3):294–295. 10.1111/1753-6405.12484
5. El-Sayed AAI, Alsenany SA, Almalki RSE, Asal MGR. Fostering creativity-nurturing behaviors among nurse educators: Investigating the interplay between evidence-based practice climate and artificial intelligence competence self-efficacy. *Nurse Educ Today*. 2025;151:106734. 10.1016/j.nedt.2025.106734.
6. Peng X, Yang L, Yang S, Wang X, Song W. Effects of clinical learning environment on the professional identity of Chinese nursing students: multiple mediating effects of learning engagement and stress-related growth. *BMC Nurs*. 2025;24(1):1387. 10.1186/s12912-025-03641-5.
7. Chen W, Zhu X, Sun S, Liao S, Guo Z. The Impact of Employees' Psychological Capital on Innovative Work Behavior: The Chain Mediating Effect of Knowledge Donating and Knowledge Collecting. *Front Psychol*. 2021;12:761399. 10.3389/fpsyg.2021.761399.
8. Zhang X, Han J, Ding R, et al. Research progress on innovative literacy assessment tools for nursing college students. *Occup Health*. 2024;40(17):2439–43. 10.13329/j.cnki.zyyjk.2024.0422.
9. Wallach MA, Torrance EP. Torrance Tests of Creative Thinking: Norms – Technical Manual. *Am Educ Res J*. 1968;5(2):272. 10.2307/1161826.
10. Sousa VD, Rojjanasirrat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract*. 2011;17(2):268–74. 10.1111/j.1365-2753.2010.01434.x.
11. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–49. 10.1007/s11136-010-9606-8.
12. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42. 10.1016/j.jclinepi.2006.03.012.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. 10.1136/bmj.n71.
14. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev*. 2007;2007(2):MR000010. 10.1002/14651858.MR000010.pub3.
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. 10.1371/journal.pmed.1000097.
16. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171–9. 10.1007/s11136-017-1765-4.
17. Elsmann EBM, Mokkink LB, Terwee CB, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Qual Life Res*. 2024;33(8):2029–46. 10.1007/s11136-024-03634-y.
18. Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic Reviews. A Product from the ESRC Methods Programme. *Lanc Univ*. 2006. 10.13140/2.1.1018.4643.
19. Wang M, Lin X. Williams' Workbook on Creative Thinking Activities. *Bull Spec Educ*. 1986;2:231–50. https://tci.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dncresource&s=id=%22A.87021192%22.&searchmode=basic&tcihsspage=tcisearch_opt1_search.
20. Warner J. Creativity and innovation effectiveness profile. Amherst Massachusetts:HRD Press. 2004:13. <https://www.kdocs.cn/l/cg7B3eg82tjU>
21. Ye S, Zhao X, Hu Z, et al. Adaptation and validation of Chinese version of the creativity and innovation effectiveness profile. *BMC Nurs*. 2025;24(1):1087. 10.1186/s12912-025-03746-x.
22. Wang QQ. Research on Creativity Characteristics of Undergraduates and Exploratory Study on Creativity Consciousness Cultivation of Undergraduates. *Chongqing University*. 2012. <https://www.kdocs.cn/l/csOc31ocBqdc>
23. Bao L, Wang L, Zhang YQ. Development and analysis of reliability and validity of Nurse Innovative Behavior Scale. *J Shanghai Jiaotong Univ (Med Sci)*. 2012;32(08):1079–82. 10.3969/j.issn.1674-8115.2012.08.025.
24. Jiang LP. A Study on the Relationships among University Innovation Climate, College Students' Innovation Self-Efficacy, and Innovation Behavior. *Qufu Normal University*. 2018. <https://www.kdocs.cn/l/cnbzwKG2eUzP>
25. Xin TT. Research on Affecting Factors of College Students' Innovation Behaviors. *Harbin Inst Technol*. 2016. <https://www.kdocs.cn/l/cqZeZz9tCG9W>
26. Zhao ZQ, Chen XH, Liu YJ, et al. Study on cultivation of nursing students' innovation behaviors based on constructivism theory teaching. *Chin Nurs Res*. 2012;26(8):752–3. 10.3969/j.issn.1009-6493.2012.08.045.
27. Shen YJ, Liu YB. Compilation of innovation and entrepreneurship status scale for nursing undergraduates. *Med High Vocat Educ Mod Nurs*. 2021;4(01):12–5. 10.3969/j.issn.2096-501X.2021.01.004.
28. Wang J, Zhao X, Song SM, et al. Current situation and influencing factors of innovation ability of master's degree nursing students. *J Nurs Adm*. 2014;14(9):609–11. <https://www.kdocs.cn/l/cdyENzYrjpTO>.
29. Scott SG, Bruce RA. Determinants of Innovative Behavior: A Path Model of Individual Innovation in the Workplace. *Acad Manage J*. 1994;37(3):580–607. 10.2307/256701.
30. Janssen O. Job demands, perceptions of effort-reward fairness and innovative work behaviour. *J Occup Organ Psychol*. 2000;73(3):287–302. 10.1348/096317900167038.
31. Weerasekera I, Hall M, Shaw L, Kiegaldie D. Instruments evaluating the quality of the clinical learning environment in nursing education: An updated systematic review. *Nurse Educ Pract*. 2023;71:103732. 10.1016/j.nepr.2023.103732.

32. Niederberger M, Schifano J, Deckert S, et al. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. PLoS ONE. 2024;19(8):e0304651. 10.1371/journal.pone.0304651.
33. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147–57. 10.1007/s11136-018-1798-3.
34. Bandura A, Freeman WH, Lightsey R, Self-Efficacy. The Exercise of Control. J Cogn Psychother. 1999;13(2):158–66. 10.1891/0889-8391.13.2.158.
35. Charette M, McKenna LG, Maheu-Cadotte MA, Deschênes MF, Ha L, Merisier S. Measurement properties of scales assessing new graduate nurses' clinical competence: A systematic review of psychometric properties. Int J Nurs Stud. 2020;110:103734. 10.1016/j.ijnurstu.2020.103734.
36. Gillespie M, Peterson BL. Helping novice nurses make effective clinical decisions: the situated clinical decision-making framework. Nurs Educ Perspect. 2009;30(3):164–70. <https://www.kdocs.cn/l/cmm0IH1ZJbHZ>.
37. Anderson JA, Willson P. Knowledge management: organizing nursing care knowledge. Crit Care Nurs Q. 2009;32(1):1–9. 10.1097/01.CNQ.0000343127.04448.13.
38. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA. 2002;287(2):226–35. 10.1001/jama.287.2.226.
39. Flake JK, Pek J, Hehman E. Construct Validation in Social and Personality Research. Social Soc Psychol Personal Sci. 2017;8(4):370–8. 10.1177/1948550617693063.
40. Bourne PA. The validity of using self-reported illness to measure objective health. N Am J Med Sci. 2009;1(5):232–8. 10.4297/najms.2009.5232.

Figures

Assessment Tools for Innovation Ability Among Nursing Students: A Systematic Review

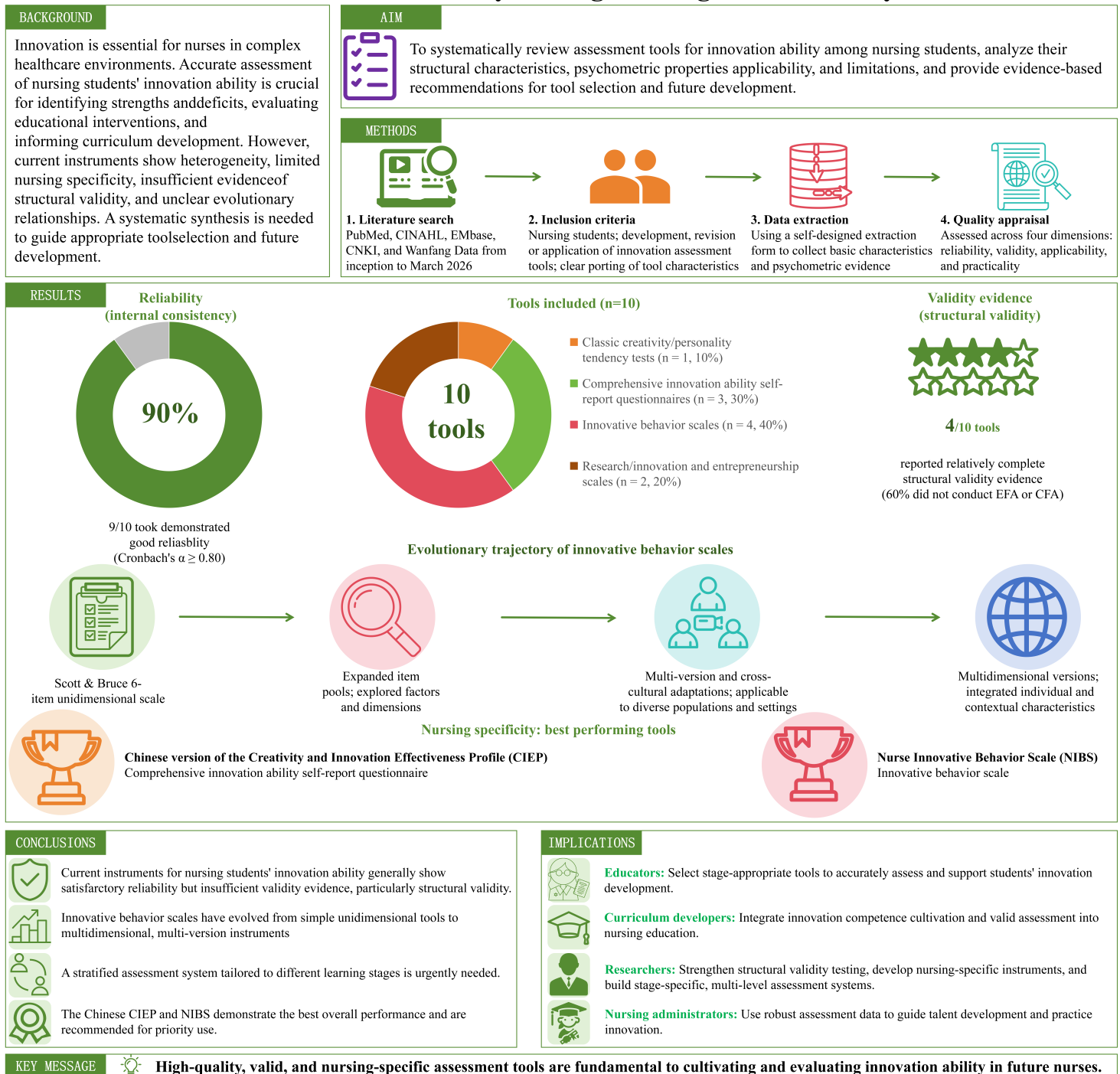


Figure 1

Graphical Abstract.

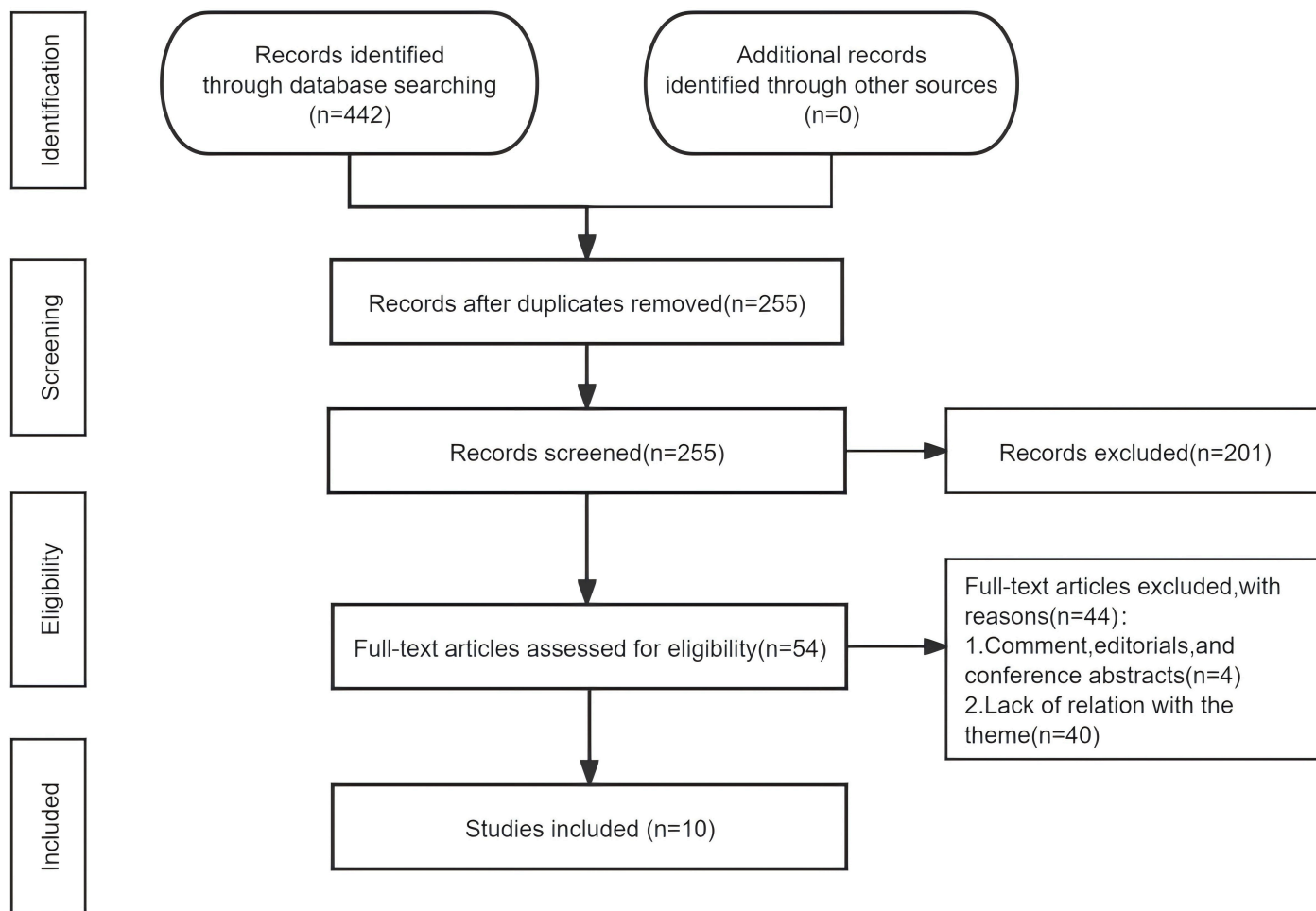


Figure 2

Flow chart of the selection process.

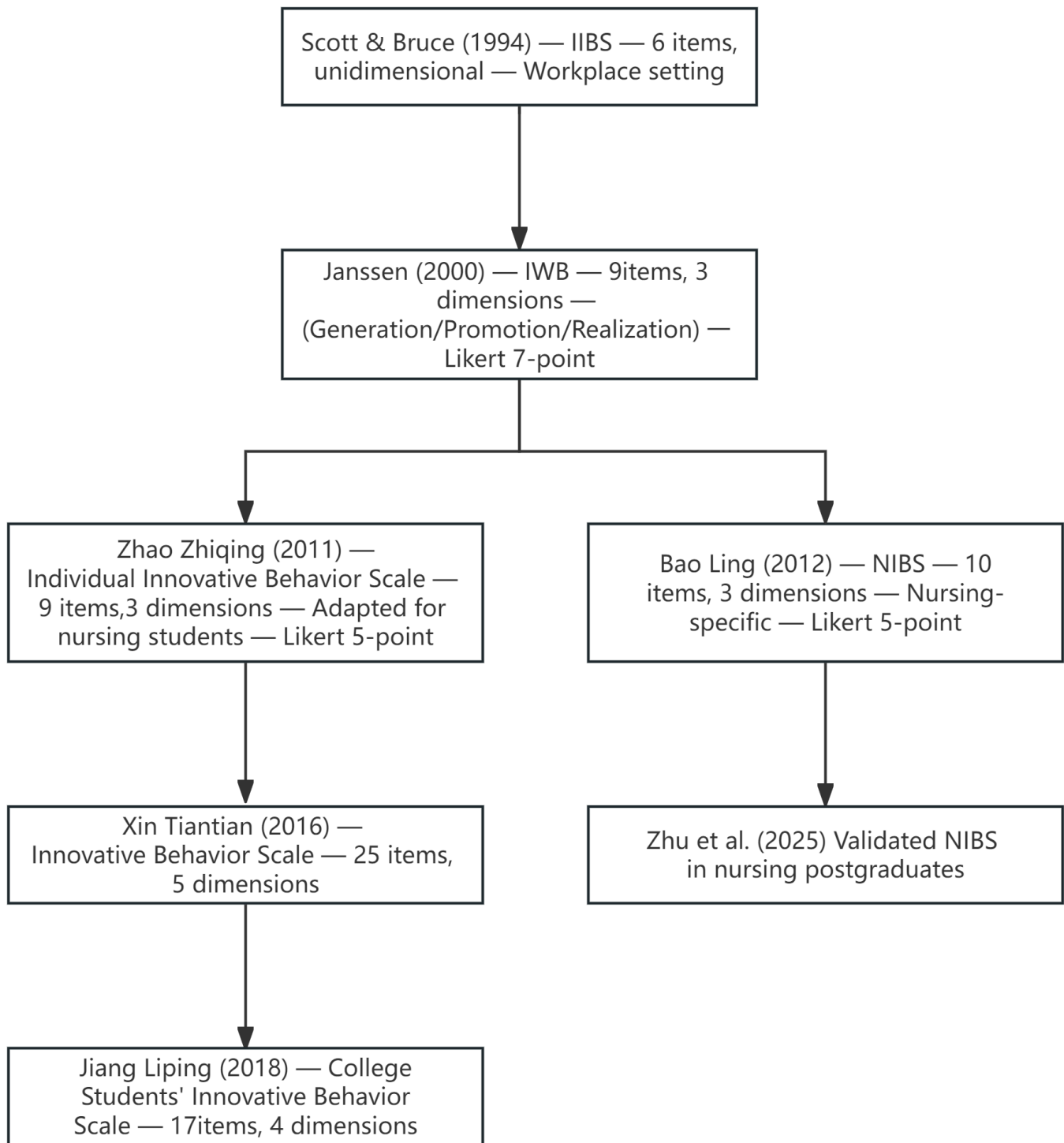


Figure 3

Evolutionary trajectory of the innovative behavior scales (1994–2018).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.docx](#)
- [Table2.docx](#)
- [SupplementaryMaterials.docx](#)