Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Psychometric properties of instruments used to measure the competence of operating room nurses: A COSMIN-based systematic review

Xinxin Xu
West China Hospital of Sichuan University
Yuanyuan Mou
Sichuan University
Shiyu Zhuo
Sichuan University
Weizhen Bai
West China Hospital of Sichuan University
Lingli Li
lilingli2000@126.com
West China Hospital of Sichuan University

Research Article

Keywords: Instruments, Systematic review, Measurement properties, Operating room nurse, competence, Evaluation tools

Posted Date: March 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8671837/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

Objective:This systematic review aimed to evaluate competence instruments for operating room nurses,in order to provide evidence-based support for selecting the most appropriate instrument for specific populations.

Background:The operating room nurses' competence significantly affects the quality of nursing care, as well as the outcomes of surgical treatments and patient prognosis.Despite the development of various instruments designed to assess the operating room nurses' competence, a comprehensive and rigorous systematic review that provides an overview and critical appraisal of these instruments is currently lacking. Consequently,a systematic review of the measurement properties of such instruments is essential.

Design: Systematic review of measurement properties following the COSMIN guidelines.

Methods：Relevant studies were searched on seven English databases and four Chinese databases from inception to October 2025.The eligible studies were reviewed to identify the competence measured by the instruments for operating room nurses. The psychometric properties of instruments in development and psychometric testing design studies were also examined. The COSMIN checklist was used to assess the studies' methodological quality.

Results:A total of 16 instruments were included,and included instruments assessed only two to four psychometric properties.No studies report measurement error and responsiveness.The PPCS-R and the SPLINTS were rated as A recommendations, while the remaining instruments were rated as B recommendations.

Conclusion: The study identified the PPCS-R and the SPLINTS as appropriate measures for assessing the competence of operating room nurses.Future research should adhere to the COSMIN methodological standards to strengthen the development and validation of competence assessment tools.

1. Introduction

Quality of care and patient safety are arguably the dominant themes of the modern healthcare agenda.Approximately 12% of preventable patient harm causes permanent disability or patient death and is mostly related to therapeutic management,and invasive clinical procedures[1].Enhance nursing competencies in therapeutic management and procedural safety.Decrease in preventable harm,disability and death,improving overall patient safety[2, 3].Competence has been described in various ways; as a combination of specific knowledge, skills, concepts, and values to meet performance standards in a particular context[4, 5].After the concept of competence was introduced into the nursing field, the International Council of Nurses (ICN) defined the competence of the generalist nurse as the knowledge, skills, judgment, and personal attributes required for nurses to fulfill their job responsibilities[6].The cultivation of competence is an important topic in the development of global nursing education.The competence of operating room(OR) nurses significantly affects the quality of nursing care,as well as the outcomes of surgical treatments and patient prognosis[7, 8].

The Association of PeriOperative Registered Nurses (AORN) established evaluation criteria for the competence of operating room nurses at the end of the 20th century, providing reference for other countries around the world[9]. At the beginning of the 21st century, the Australian College of Operating Room Nurses (ACORN) focused on the requirements for the competence of scrub nurses.The competence of OR nurses, also known as perioperative nurses,were divided into technical skills and non-technical skills. Technical skills (TS) include theoretical practical knowledge, mastery of aseptic techniques, and strict adherence to various operational procedures, while non-technical skills (NTS), also known as soft skills, mainly include holistic nursing, situation awareness, empathy, collaboration, communication, task management, and teamwork[10, 11]. NTS have significant individual differences but are crucial in OR nursing work. The lack of NTS can lead to interruptions in surgery, retained surgical objects, and other adverse events, threatening the safety of surgical patients and even endangering their lives[11].

Despite the existence of several instruments designed to assess the competence of OR nurses, these tools have been developed in disparate contexts.Consequently, there is no consensus on which tool is the most appropriate and reliable, which has led to confusion among researchers and OR managers when selecting it.To our knowledge,there is no systematic review to summarize all the OR nurses' competence assessment tools and report their psychometric properties.Therefore, this study adopt the COSMIN guideline to evaluate instruments,and adopt the GRADE guideline to form recommendations.This study evaluated OR nurses' competence instruments using the COSMIN guidelines,and aim to provide an evidence-based assessment tool for the competence of OR nursing teams and lay the foundation for future research.

2. Method

This study used the COSMIN guideline for quality assessment of included studies[12, 13],and was registered on PROSPERO(CRD42024627674)

2.1Search strategy

We conducted primary searches in PubMed using both MeSH terms and free terms to develop search words.The identified search strategy was confirmed by our research group.This study systematically searched PubMed,Embase,Web of Science,CINAHL,PsycInfo,China National Knowledge Infrastructure(CNKI) and Wanfang Database using the identified search strategy, with the search period from database inception to October 2025.And manually trace the references and related studies to supplement this research.Our search included keywords such as"operating room nurs*,""theatre nurs*,""competenc*,""perioperative nurs*,""scrub nurs*,""circulating nurs*,""capacit*,""abilit*,""questionnaire*," "instrument*," "tool*,""assess*," "measure*"and"scale*." The search strategies provided in Supplemental Appendix A, NoteExpress was used to export and manage the retrieved studies.

2.2. Inclusion and exclusion criteria

The inclusion criteria were as follow: (1) the study was conducted with OR nurses, including circulating nurses and scrub nurses; (2) Research involving the development or validation of measurement instrument of OR nurses' competence and evaluate one or more psychometric properties;and(3)studies published in English or Chinese.Exclusion criteria for this study were as follow: (1) unavailability of the full text; (2)non-English or non-Chinese literature; (3)duplicate publications.

2.3Literature screening and data extraction

According to the inclusion and exclusion criteria, two researchers trained in evidence-based methodology independently screened the literature and extracted data,If a consensus could not be reached,a third researcher was consulted to resolve the disagreement.All studies were imported into the NoteExpress software, where they were processed through a series of deduplication, initial screening, and full-text screening steps.The research team determined whether to include them through group discussions.s. We extracted data on (1) Characteristics of the included studies.(Table 1); (2)Methodological quality and measurement properties evaluation of the included instruments(Table 2); and(3)Summary of results for psychometric properties and evidence(Table 3)

2.4.Assessment of methodological quality

COSMIN risk of bias checklist is a consensus-based standard for the selection of health measurement instruments,which can evaluate the methodological quality and measurement properties,and provide recommendations for instrument selection[12, 13].The COSMIN risk of bias checklist has 10 domains and 116 items, including patient-reported outcome measure(PROM) development, content validity, structural validity, internal consistency,cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity,and responsiveness. Each item has five options, including "very good","adequate", "doubtful", "inadequate", and "not applicable". The "worst score counts" principle was used to determine the overall quality of the study[12].

2.5.Measurement properties quality assessment

With the COSMIN criteria for good measurement properties,Content validity and other measurement properties could be rated as sufficient(+),insufficient(-),or indeterminate(?).

2.6.Evidence Level Assessment

The modified Grading of Recommendations Assessment,Development,and Evaluation (GRADE) system [12, 13]is used to grade the quality of evidence; the evaluation tools are downgraded based on four criteria: risk of bias, inconsistency, indirectness, and imprecision, with evidence levels being "high," "moderate," "low," or "very low";Content validity is "sufficient" (any level of evidence),with adequate internal consistency (at least low-quality evidence) was rated as Category A,which is recommended for use.Its measurement attributes are 'inadequate'(at high-quality evidence) was rated as Category C,which is not recommended for use.All others were not in Category A or Category C rated as Category B,which has potential for application, but further research is required to evaluate its quality.

3. Results

3.1. Literature search

As shown in Fig. 1.,a total of 11154 articles were initially retrieved from the seven databases searching, of which 9527 remained after the removal of duplicates. Subsequently, 9412 articles were excluded after review of their titles and abstracts, and an additional 115 articles were excluded after full-text review. Ultimately, a total of 16 articles were included in the present study after screening for eligibility.The literature screening flowchart is shown in Fig. 1,The basic characteristics of the included literature are shown in Table 1.This study included a total of 8 evaluation tools, consisting of 4 comprehensive competence evaluation tools and 4 specialized competence evaluation tools.

Table 1
Characteristics of the included studies.

Reference	item	Instrument name	Developer/year	country	Original language	Theoretical model	Population(N)	Content and/or subscales (items
[14]	comprehensive competency evaluation tools	The core competency scale for ORNs(CCS_ORN)	Na Mei/2022	China	Chinese	—	300	1.Emergency and problem-solving skills(5 items) 2.Specialist knowledge and skills (12 items) 3.Collaboration a interpersonal communication skills (5 items) 4.Teaching and research capacit (8 items) 5.Laws and regulations a relevant expertis (5 items) 6.Management capabilities and basic skills (7 items)
[15]		the Revised Perioperative Competence Scale(PPCS-R)	Brigid M. Gillespie/2012	Australia	English	—	1122	1.Foundational Knowledge & Skills(9 items) 2.Leadership(8 items) 3.Collaboration(items) 4.Proficiency(6 items) 5.Empath items) 6.Professional Development(6 items)
[16]		The PPCS-R in the United States context	2024	the United States	English	—	1581	1.Foundational Knowledge and Skills(9 items) 2.Leadership(8 items) 3.Collaboration(items) 4.Proficiency(6 items) 5.Empath items) 6.Professional Development(6 items)
[17]		the PPCS-R in the Swedish context	2018	Swiss	Swedish	—	1033	1.Foundational Knowledge and Skills(9 items) 2.Leadership(8 items) 3.Collaboration(items) 4.Proficiency(6 items) 5.Empath items) 6.Professional Development(6 items)
[18]		the Persian PPCS-R	Neda Mirbagher Ajorpaz/2017	Iran	Persian	—	357	1.Foundational Skills and Knowledge (nine items) 2.Leadership (eig items) 3.Collaboration (items), 4.Proficiency (six items) 5.Empath (five items) 6.Professional Development (si items)
[19]		the Turkish version PPCS	2019	Turky	Turkish	—	206	1.foundational knowledge and skills(9 items) 2.leadership(8 items)

Reference	item	Instrument name	Developer/year	country	Original language	Theoretical model	Population(N)	Content and/or subscales (items
								3.collaboration(6 items) 4.proficie (6 items) 5.empathy(5 iter 6.professional development(6 items)
[20]		The Chinese Perceived Perioperative Competence Scale-Revised (C-PPCS-R)	2022	China	Chinese	—	480	1.Foundational Skills and Knowledge (8 items) 2.Leadership (7 items) 3.Collaboration (items), 4.Proficiency (4 items) 5.Empath (5 items) 6.Professional Development (5 items)
[21]		the Perceived Perioperative Competence Scale-Revised Short Form(PPCS-R-SF)	2023	Australia	English	—	649	1.Foundational Skills and Knowledge (3 items) 2.Leadership (3 items) 3.Collaboration (items), 4.Proficiency (3 items) 5.Empath (3 items) 6.Professional Development (3 items)
[22]		the Perceived Perioperative Competence Scale–Revised instrumen into Italian(PPCS-R ITA)	2025	Italy	Italian	-	509	1.Foundational Skills and Knowledge(9 items) 2.Leadership(8 items) 3.Empath items) 4.Professional Development(6 items) 5.Communicatio items) 6.Proficiency(6 items)
[23]		Competency Evaluation Index System for Nurses in Operating Room	Wang Yumei/2015	China	Chinese	Icebery model,onion model	60	1.specialized knowledge(8 iter 2.professional ability(5items) 3.personality(8 items) 4.self-motivation(3 iter
[24]		Operating Room Professional Nurses Core Competency Evaluation Scale	Mu Li/2019	China	Chinese	—	433	1.professional development ability(12 items) 2.professional cooperation ability(12 items) 3.professional management ab (10 items) 4.perioperative assessment abil (6 items)5.basic professional ability(5 items).
[25]	specialized competency evaluation tools	the Imperial College Assessment of Technical Skill for Nurses (ICATS-N)	Nick Sevdalis/2009	United Kingdom	English	—	21	1.Gowning and Cloving(3 items) 2.Setting up instrumentation(items) 3.Draping items) 4.Maintaining sterile field(4 items)

Reference	item	Instrument name	Developer/year	country	Original language	Theoretical model	Population(N)	Content and/or subscales (items
[26]		the Scrub Practitioners' List of Intraoperative Non-Technical Skills(SPLINTS)	Lucy Mitchel/2011	United Kingdom	English	—	34	1.Situation awareness(3 iter 2.Communicatio and teamwork(3 items)3.ask management(3 items)
[27]		The Chinese Version of Scrub Practitioners' List of Intraoperative Non-Technical Skills(C-SPLINTS)	2021	China	Chinese	□	118	1.Situation awareness(3 iter 2.Communicatio and teamwork(3 items)3.ask management(3 items)
[28]		the Non-technical Skills for Operating Room Nurses behavioural marker system (NOTSORN)	Irene Sirevåg/2024	Norway	Norwegian and English	—	25	1.Professional accountability(2 items) 2. Situational awareness and decision making(4items) Teamwork and communication(items) 4. Leadership and task management(3 items)
[29]		The circulating practitioner's list of non-technicalskills(CPLNTS)	Reza Kalantari/2021	Iran	English	—	316	1.task management(13 items) 2.teamwork(7 items) 3.situatio awareness(7 iter 4.communicatio Items) 5.leadership(7 items)

"—"Indicates that there is no report.

3.2.Study description

The characteristics of the included studies are delineated in Table 1.A total of 7,244 study subjects were included, with sample sizes ranging from 21 to 1,581.Among the included studies, only three were published in Chinese, with the remainder in English.This table lists nine original tools and seven culturally adapted tools.The Revised Perioperative Competence Scale (PPCS-R) is the most widely culturally adapted questionnaire, having been validated in six countries.[16–19, 22, 30].All research constructs were the competence of OR nurses,including ST and NST.The original and translated versions of a category of instruments were evaluated individually as separate instruments due to the variations in their items and sub-instrument compositions.

3.3. Quality assessment

3.3.1 Methodological quality assessment

As presented in Table 2, of the 16 included studies, only two studies explored more than five psychometric properties[15, 31]. However, there was insufficient information on cross-cultural validity,measurement error and responsiveness. Content validity were consistently reported across all studies.

3.3.2 Quality of measurement properties

The quality of PROM development was assessed as "adequate" in one studies[15, 28],which had representative sample of the target population.A total of fifteen studies rated the PROM development study as 'doubtful,' of which ten conducted cognitive interviews with professionals or researchers only[14, 15, 17, 19, 23–25, 27–29];three conducted cognitive interviews or pre-tests with the target population, but it was unclear whether they asked about comprehensibility and comprehensiveness[16, 22, 30]□two did not describe the research process in detail[18, 31].Content validity was evaluated through cognitive interviews or pilot tests.The quality of content validity was rated as "adequate" in three studies[15, 26, 28],where both target population and professionals were consulted regarding the relevance and comprehensiveness of the constructs;these studies employed a sufficiently representative target population and professionals and used appropriate methods to analyze the data.Furthermore,thirteen studies did not provide detailed descriptions of their research designs or data analysis processes for content validity, resulting in the methodological quality of content validity being rated as 'doubtful.'

The structural validity of the instruments was assessed across 13 studies.In this studies,confirmatory factor analysis (CFA) is the most frequently used method.Two studies employed exploratory factor analysis (EFA).[14, 15, 22, 24, 27],six studies used both CFA and EFA[14, 15, 17, 27, 29, 30],and only one applied item response theory[16].Five studies which used both CFA and EFA with adequate sample sizes and no other significant methodological flaws, so

the methodological quality was assessed as ‘very good’,the Chinese version of SPLINTS had a sample size of less than five times the required amount and was rated as ‘inadequate’[27].The PPCS-R in the United States context used item response theory demonstrated sufficient model and sample size, thus resulting in an assessment of its methodological quality as ‘very good’[16].Four studies received an ‘doubful’ rating, primarily due to their limited focus on EFA and inadequate sample size[18, 19, 22, 24].The remaining three studies were characterised by adequate sample sizes and the absence of any other significant methodological weaknesses,consequently they were rated as 'adequate'.Internal consistency was evaluated except NOTSORN,and this study report Cronbach’s alpha,and the quality was consistently rated as ‘very good’.However,four studies did not report Cronbach’s alpha or item-total correlations calculated,and so were consequently rated as ‘doubful’[22, 25, 26, 29].Only one study assess measurement invariance by categorizing perioperative nurses into participants with perioperative nursing experience,used multi-group confirmatory factor analysis framework,and did not have other important methodological flaws,so it was rated as ‘very good’.[16]

Reliability was assessed in eight studies, which calculated intragroup correlation coefficients (ICC) or Pearson's correlation.Nonetheless, these studies were rated as 'doubful' ,because insufficient information of detailed formula descriptions and information about whether the target population was stable or not and whether the measurement conditions were similar[14, 15, 18, 19, 23, 25, 29, 31].The criterion validity of the PPCS-R-SF was based on other commonly used scales and rated as ‘doubful’[31].In terms of hypothesis testing, one study adequately described most of the important characteristics of the known-groups,had no other methodological flaws and was therefore rated as ‘adequate’[15].

The criterion validity of the PPCS-R-SF was based on the previously validated 40-item PPCS-R and meet the COSMIN guideline criteria,therefore it was rated as ‘very good.’[31]In terms of hypothesis testing,one study adequately described most of the important characteristics of the subgroup and had no other methodological flaws,PPCS-R was rated as ‘adequate.’[15]

3.3.2 Quality of measurement properties

Table 2 shows the quality of psychometric properties retrieved from included studies.Content validity was assessed for all instruments in sixteen studies, and three instruments were rated as ‘sufficient (+)’[15, 23, 26, 28].Thirteen instruments received an ‘inconsistent (±)’rating[14, 16–20, 22–24, 29–32].Structural validity was assessed for thirteen instruments in sixteen studies, with six instruments received a ‘sufficient (+)’ rating[15, 17, 22, 26, 29, 30],and four were rated as ‘insufficient (-)’[16, 19, 21, 29] and three instruments received an ‘indeterminate (?)’ rating[14, 18, 24].Internal consistency was assessed for fifteen instruments in sixteen studies,with seven instruments rated as ‘sufficient (+)’[15, 16, 19, 22, 26, 30, 31],and seven studies were assigned a rating of ‘indeterminate (?)’[17, 23–27, 29, 33, 34] and only one study rated as ‘indeterminate (?)[18]. Reliability was assessed for eight instruments in sixteen studies, with four instruments rated as ‘sufficient (+)’[18, 19, 29, 31],three instruments rated as ‘indeterminate (?)’[15, 23, 25],and one instrument rated as ‘sufficient (+)’[14].Only one study assessed measurement invariance,and PPCS-R in the United States context was rated as ‘sufficient (+)’[16].The criterion validity was rated as ‘sufficient(+)’ for PPCS-R-SF[31],and hypotheses testing for construct validity was rated as ‘indeterminate (?)’ for PPCS-R[15].

Table 2
Methodological quality and measurement properties evaluation of the included instruments.

Instrument	PROM development	Content validity		Structural validity		Internal consistency		Reliability		Measurement invariance		Criterion validity		Hypotheses testing for construct validity	
	MQ	MQ	QR	MQ	QR	MQ	QR	MQ	QR	MQ	QR	MQ	QR	MQ	QR
CCS_ORN[14]	D	D	?	V	?	V	?	A	-						
PPCS-R[15]	A	A	+	V	+	V	+	D	?					A	?
PPCS-R in the United States context[16]	D	D	?	V	-	V	+			V	+				
Persian PPCS-R[18]	I	D	?	D	?	V	-	A	+						
PPCS-R in the Swedish context[17]	D	D	?	V	+	V	?								
Turkish PPCS-R[19]	D	D	?	D	-	V	+	A	+						
C-PPCS-R[30]	D	D	?	V	+	V	+								
PPCS-R-SF[31]	D	D	?	A	-	V	+	D	+			V	+		
PPCS-R ITA[22]	D	D	?	D	+	D	+								
Competency Evaluation Index System for Nurses in Operating Room[23]	D	D	?			V	?	V	?						
Operating Room Professional Nurses Core Competency Evaluation Scale[24]	D	D	?	A	?	V	?								
ICATS-N[25]	D	D	?			D	?	D	?						
SPLINTS[26]	D	A	+	A	+	D	+								
C-SPLINTS[27]	D	D	?	I	-	V	?								
NOTSORN[28]	A	A	+												
CPLNTS[29]	D	D	?	V	+	D	?	A	+						

Comments: MQ, Methodological quality; QR, Quality of he results rated,V: very good, A: adequate, D: doubtful, I: inadequate; "+ ": sufficient; "-": insufficient, "?". .Empty cells indicate that the measurement property was not assessed in the study

3.3.3. Evidence grading and recommendations

Table 3 summarizes the evidence quality of all PROMs, rated as 'high,' 'moderate,' 'low,' or 'very low' based on downgrading factors including risk of bias, inconsistency, imprecision and indirectness.None of the PROMs were downgraded for inconsistency, imprecision, or indirectness.For risk of bias,the PPCS-R,the SPLINTS and the NOTSORN were rated 'adequate'[15, 28, 35],so they were not downgraded,while the rest instruments were downgraded one level due to the methodological quality of 'doubtful' for content validity.The methodological quality rating for the structural validity of the Persian PPCS-R,the Turkish PPCS-R and the PPCS-R ITA were 'doubtful'[18, 19, 22], so they was downgraded by one level.The C-SPLINTS was rated 'inadequate'[27],so it was downgraded by two level,while the rest of the instruments were rated 'adequate' or 'very good', so they were not downgraded. The methodological quality of the internal consistency of the PPCS-R ITA,the ICATS-N,the SPLINTS and the CPLNTS was 'doubful' and the above four instruments were downgraded one level[22, 25, 29, 35]. The rest of the instruments were not downgraded, due to the 'adequate' or 'very good' rating. The methodological quality of the reliability of the PPCS-R and the PPCS-R-SF was rated as 'doubtful' and were therefore downgraded one level[15, 31]. The remaining instruments are rated 'adequate' or 'very good' and do not require downgrading. The methodological quality of the measurement invariance of the PPCS-R in the United States context[16], the criterion validity of the PPCS-R-SF[31],and the hypotheses testing for construct validity of the PPCS-R[15] were 'adequate' or 'very good' and they were not downgraded.For inconsistency, we considered each different language versions of a PROM a unique PROM. In terms of imprecision, the ICATS-N reporting reliability had retest sample sizes of less 50 and were therefore downgraded two level[25].AND sample sizes of 50–100 of the competence Evaluation Index System for Nurses in OR,therefore downgraded one level[23]. None of the studies were indirect, so no downgrading.

The recommended grades for the sixteen instruments are presented in Table 3. According to the COSMIN guidelines, the content

validity and internal consistency of the PPCS-R and the SPLINTS are 'sufficient'[15, 35],and the level of evidence is no less than 'moderate',making it an A recommendation[15, 35].Furthermore, none of the PROMs have high-level evidence indicating insufficiency(-),resulting in none of the C recommendations. The other fourteen instruments were all category B recommendations with potential for application, but further research is needed to evaluate their psychometric properties.

Table 3
Summary of results for psychometric properties and evidence.

Instrument	Content validity		Structural validity		Internal consistency		Reliability		Measurement invariance		Criterion validity		Hypotheses testing for construct validity		Recomm grade
	QR	QE	QR	QE	QR	QE	QR	QE	QR	QE	QR	QE	QR	QE	
CCS_ORN[14]	?	Moderate	?	High	?	High	-	High							B
PPCS-R[15]	+	High	+	High	+	High	?	Moderate					?	High	A
PPCS-R in the United States context[16]	?	Moderate	-	High	+	High			+	High					B
Persian PPCS-R[18]	?	Moderate	?	Moderate	-	High	+	High							B
PPCS-R in the Swedish context[17]	?	Moderate	+	High	?	High									B
Turkish PPCS-R[19]	?	Moderate	-	Moderate	+	High	+	High							B
C-PPCS-R[30]	?	Moderate	+	High	+	High									B
PPCS-R-SF[31]	?	Moderate	-	High	+	High	+	Moderate			+	High			B
PPCS-R ITA[22]	?	Moderate	+	Moderate	+	Moderate									B
Competency Evaluation Index System for Nurses in Operating Room[23]	?	Moderate			?	High	?	Low							B
Operating Room Professional Nurses Core Competency Evaluation Scale[24]	?	Moderate	?	High	?	High									B
ICATS-N[25]	?	Moderate			?	Moderate	?	Very low							B
SPLINTS[35]	+	High	+	High	+	Moderate									A
C-SPLINTS2020[27]	?	Moderate	-	Low	?	High									B
NOTSORN[28]	+	High													B
CPLNTS[29]	?	Moderate	+	High	?	Moderate	+	High							B

Comments: (1) Overall rating rated as Sufficient (+), Intermediate (?), Insufficient (−). (2) MQ:Methodological quality; QR:Quality of he results rated,(3) Empty cells indicate that the measurement property was not assessed in the study.

4. Discussion

This review evaluated 16 OR nurses' competence assessment instruments based the COSMIN guidelines. To our knowledge,our study is the first to comprehensively assess the psychometric properties of competence assessment tools for operating room nurses. This systematic review provides the first comprehensive synthesis and critical evaluation of the operating room nurse competence assessment instruments.

Content validity is the most important psychometric characteristic in the evaluation process, and evaluators should first assess content validity[12, 36]. The COSMIN guidelines evaluate content validity from the relevance, comprehensiveness, and understand ability of the study.The methodological quality OF development is predominantly rated as 'doubtful',and issues include item generation based on the opinions of professional nurses but limitations in the representativeness of the sample,lack experienced or trained interviewers, appropriate interview outlines, at least two researchers independently coding, analyzing interview data,which raises concerns about the consistency and reliability of the findings.The present study is limited by the fact that it is based solely on the findings of related studies which interviewed professional OR nurses or managers. In addition,the work of OR nurses is characterised by team cooperation with surgeons, anesthesiologists and others,and teamwork in the OR necessitates a comprehensive understanding of the competence required of OR nurses, as determined by their teamwork members rather than solely by the competencies stipulated by OR nurses themselves.In order to address the aforementioned issues, the utilisation of standardised qualitative methodologies, in conjunction with the meticulous delineation of the developmental process and content evaluation procedure of PREMs, has the potential to augment the scientific rigour of these instruments.The comprehensive competence assessment tools for OR nurses vary in evaluation dimensions and content, but the dimensions of evaluation mostly revolve around the ICN nurse

competence theoretical framework.The ICN defines nurse competence as the knowledge, skills, judgment, and personal qualities nurses need to provide safe and ethical services to patients[37]; Specialized competence assessment tools predominantly concentrate on the TS and NTS of scrub nurses, the NTS of circulating nurses and both.It is evident that research on non-technical skills is gradually increasing, and researchers have found that NTS have a significant impact on the safety of surgical patients and the quality of surgical procedures[11].Nevertheless, the available assessment tools remain inadequate and require further development and refinement.

Structural validity refers to the degree of fit between the scale dimensions and the dimensions of the construct being measured, and is also a prerequisite for explaining internal consistency analysis[13].Structural validity preferred statistical methods based on item response theory(IRT) analyses,but only the PPCS-R in the United States context used IRT,The COSMIN guidelines prescribe the use of factor analysis for the appraisal of structural validity, considering that CFA is better than EFA.CFA is suitable for situations where the dimensions of the construct being measured are determined, while EFA is suitable for situations where the dimensions of the construct being measured are uncertain. The evaluation tools included in this study mainly applied EFA,and eight studies implemented CFA, this limitation has affected the evaluation of the structural validity of the study. It is recommended that in future development studies of OR nurse competence assessment tools, confirmatory factor analysis should be adopted, and the sample size should be ensured to be sufficient to improve the structural validity of the assessment tools and the methodological quality of their internal consistency.Internal consistency was reported for most instruments except NOTSORN in this study. However, the authors noted that many studies relied solely on Cronbach alpha to assess reliability, which is not sufficient. Just PPCS-R in the United States context,Persian PPCS-R and SPLINTS used other internal consistency indices like Omega coefficient, Pearson’s correlation coefficient.This multidimensional approach facilitates a more comprehensive evaluation of the measurement instrument's reliability, thereby enhancing the credibility of the study results. According to the COSMIN guidelines, there is no gold standard for PROM instruments in principle, except when the original scale is simplified, where case the original scale can be used as the gold standard for newly developed brief scales. the PPCS-R-SF criterion validity study used the PPCS-R as the gold standard, which meet the COSMIN guidelines.criterion validity was assessed through Spearman’s rank correlation analysis ,which was conducted to assess the relationships between the PPCS-R or PPCS-R-SF scores and external variables,like self-efficacy, years of experience in the OR, postgraduate education and previous perioperative speciality training received.Regarding reliability, studies have not clarified whether OR nurses are stable between repeated measurement intervals and whether measurement conditions are similar.Consequently, it is recommended that subsequent studies are appraised in accordance with the COSMIN guidelines, thereby ensuring the integrity and comprehensiveness of the study design and analytic process.

This review can inform the selection of appropriate instruments to measure the competence in OR nursing,and we systematically reviewed the 16 assessment instruments following COSMIN guidelines.According to GRADE,the instruments that were incorporated were subsequently categorized into various levels of recommendation..The PPCS-R and the SPLINTS had sufficient (+) ratings.Evidence synthesis identified two strongly recommended: the PPCS-R and the SPLINTS is grade A, and they are recommended for use.The remaining instruments possess the potential for application; however, further research is necessary to assess their quality.

The PPCS-R found two groups of competencies in the PPCS-R scale: technical and nontechnical competencies.Technical competencies are represented by the domains Foundational Knowledge and Skills and Proficiency, while nontechnical competencies include Leadership, Collaboration/Collegiality, Empathy, and Professional Development.As a self-assessment tool, the PPCS-R scale focuses on the assessment of the competencies perceived by OR nurses, emphasizing how the perception of competence, which is an integral part of the self-professional image, directly and indirectly influences the role, care performance, job satisfaction, and maintenance of competencies[38].This scale is the first internationally widely used assessment tool for evaluating perioperative competence.By considering each version and each subscale of a PROM a unique PROM, the ones with the highest quality can be recommended for use in the future[13],so it is useful to evaluate the measurement properties per different language versions PPCS-R.It has been translated into Arabic, Chinese, Spanish, Swedish, and Turkish, and has conducted cross-sectional surveys in four countries[38],besides it has developed a brief version of the scale for clinical research.The SPLINTS offer structured observation, rating and feedback on the OR nurses' behaviors during surgery and aims to accelerate the development of NTS in OR nurses ,and is customized for scrub nurses.It has three categories of NTS: situation awareness, communication and teamwork, and task management, and each category has three elements concretizing the NTS. In addition,as an observational instrument, it provided examples of good and poor practice[26].However, the existing evidence on the measurement properties of these two scales is inconsistent, with significant limitations regarding sample size, cross-cultural validity and responsiveness. These methodological constraints reflect the wider challenge of measuring the competence of OR nurses, and highlight the need to continually generate new evidence based on COSMIN guideline, even for scales currently classified as highly recommended.Future research on competence for OR nurses should advance at both the conceptual and methodological levels simultaneously.Future studies should clarify the content and value of OR nurses’ competence, particularly NTS,which builds a solid theoretical foundation that is crucial for perioperative nurses' future education.Methodologically, future studies should closely follow the COSMIN recommendations for designing research and thoroughly testing psychometric validity.Establishing a measuring system based on strong theoretical foundations and rigorous methodology is a vital step in moving research on OR nurse competence forward.

The findings of our study have important implications for nursing education, particularly in the education of operating room nurses. The availability of competence assessment instruments with relatively robust psychometric evidence supports their use in competency-based education to inform curriculum design, identify learning needs, and evaluate educational outcomes.This is particularly evident in non-technical skills such as communication, teamwork, and situational awareness.These skills are imperative for ensuring patient safety,but numerous educational programs do not adequately assess them.Furthermore, the limitations of existing tools underscore the importance for nurse educators to prioritize assessment instruments developed and validated using rigorous standards, such as those recommended by COSMIN. This approach ensures meaningful and reliable evaluation of learners' competencies.

5. Limitations

This review has several limitations. First, the analysis was restricted to studies published in Chinese and English, potentially leading to the omission of competence assessment instruments for OR nurses developed and validated in other languages. Second, several of the included instruments were self-reported measures, so discrepancies between self-perceived and actual competencies may exist, which could introduce bias into the assessment results. Finally, the psychometric properties of some instruments were evaluated in only one study, which limits the robustness and generalizability of the available evidence.

6. Conclusion

This study systematically evaluated the psychometric properties of instruments used to assess competence of OR nurses in accordance with the COSMIN framework. Based on the available evidence, the PPCS-R and the SPLINTS were identified as the most appropriate instruments and were recommended for use. Future research is encouraged to be guided by the methodological framework outlined in the COSMIN Risk of Bias checklist when developing, refining, and evaluating assessment tools, in order to strengthen the quality and transparency of measurement property evidence.Robust competency assessment is essential for advancing competency-based perioperative nursing education.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 52373237)

Authors' contributions

Xin-xin Xu:Conceptualization, Writing – original draft, Supervision,Methodology, Validation.

Yuanyuan Mou:Methodology, Validation.

Shiyu Zhou:Methodology, Validation.

Weizhen Bai:Methodology, Validation.

Lin-li Li:Funding acquisition, Supervision, Writing –review & editing, Project administration.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements We would like to extend our appreciation to all members of the study team for their stimulating discussions and a supportive research environment.

Clinical trial number: not applicable.

References

1. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, Bower P, Campbell S, Haneef R, Avery AJ *et al*: **Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis**. *BMJ-BRIT MED J* 2019, **366**:l4185.
2. McHugh MD, Aiken LH, Sloane DM, Windsor C, Douglas C, Yates P: **Effects of nurse-to-patient ratio legislation on nurse staffing and patient mortality, readmissions, and length of stay: a prospective study in a panel of hospitals**. *LANCET* 2021, **397**(10288):1905-1913.
3. Aiken LH, Sloane DM, Brom HM, Todd BA, Barnes H, Cimiotti JP, Cunningham RS, McHugh MD: **Value of Nurse Practitioner Inpatient Hospital Staffing**. *MED CARE* 2021, **59**(10):857-863.
4. Prahalad CK, Hamel G: **Chapter 3 - The Core Competence of the Corporation**. In *Knowledge and Strategy*. Boston: Butterworth-Heinemann; 1999:41-59.
5. Prahalad CK, Hamel G: **The Core Competence of the Corporation**. *HARVARD BUS REV* 1990, **68**(3):79-91.

6. ALEXANDER M, RUNCIMAN P: *ICN Framework of Competencies for the Generalist Nurse: Report of the Development Process and Consultation*. Geneva: Jean-Marteall; 2003.
7. Thomasson BG, Fuller D, Mansour J, Marburger R, Pukenas E: **Efficacy of surgical safety checklist: Assessing orthopaedic surgical implant readiness**. *HEALTHCARE-J DEL SCI* 2016, **4**(4):307-311.
8. Jung JJ, Elfassy J, Juni P, Grantcharov T: **Adverse Events in the Operating Room: Definitions, Prevalence, and Characteristics. A Systematic Review**. *WORLD J SURG* 2019, **43**(10):2379-2392.
9. Conner R: **Perioperative care coordinator nurse competency statements. Nursing Practice Committee. Association of periOperative Registered Nurses**. *AORN J* 1999, **70**(4):618, 621-624, 626.
10. Falk-Brynhildsen K, Jaensson M, Gillespie BM, Nilsson U: **Swedish Operating Room Nurses and Nurse Anesthetists' Perceptions of Competence and Self-Efficacy**. *J PERIANESTH NURS* 2019, **34**(4):842-850.
11. Siu J, Maran N, Paterson-Brown S: **Observation of behavioural markers of non-technical skills in the operating room and their relationship to intra-operative incidents**. *SURG-J R COLL SURG E* 2016, **14**(3):119-128.
12. Prinsen C, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet H, Terwee CB: **COSMIN guideline for systematic reviews of patient-reported outcome measures**. *QUAL LIFE RES* 2018, **27**(5):1147-1157.
13. Mokkink LB, Elsman E, Terwee CB: **COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0**. *QUAL LIFE RES* 2024, **33**(11):2929-2939.
14. Mei N, Chang L, Zhu Z, Dong M, Zhang M, Zeng L: **Core competency scale for operating room nurses in China: Scale development, reliability and validity evaluation**. *NURS OPEN* 2022, **9**(6):2814-2825.
15. Gillespie BM, Polit DF, Hamlin L, Chaboyer W: **Developing a model of competence in the operating theatre: psychometric validation of the perceived perioperative competence scale-revised**. *INT J NURS STUD* 2012, **49**(1):90-101.
16. Maio S, Stobinski JX, Gillespie BM: **Psychometric evaluation of the United States-adapted perceived perioperative competence scale-revised: A national survey**. *INTERNATIONAL JOURNAL OF NURSING STUDIES ADVANCES* 2024, **6**.
17. Jaensson M, Falk-Brynhildsen K, Gillespie BM, Wallentin FY, Nilsson U: **Psychometric Validation of the Perceived Perioperative Competence Scale-Revised in the Swedish Context**. *J PERIANESTH NURS* 2018, **33**(4):499-511.
18. Ajorpaz NM, Tafreshi MZ, Mohtashami J, Zayeri F, Rahemi Z: **Psychometric Testing of the Persian Version of the Perceived Perioperative Competence Scale--Revised.** *J NURS MEAS* 2017, **25**(3):E162-E172.
19. Sönmez B, Ayoğlu T: **Validity and reliability of the Perceived Perioperative Competence Scale-Revised**. *NURS HEALTH SCI* 2019, **21**(4):428-435.
20. Yu Q, Wei R, Wei Y, Wu X, Liang T: **Psychometric evaluation of the perceived perioperative competence scale-revised among the Chinese operating room nurses: a methodological research.** *BMC NURS* 2022, **21**(1):79.
21. Gillespie BM, Harbeck E, Sutherland-Fraser S, Nicholson P, Boric T: **Psychometric validation of the Perceived Perioperative Competence Scale-Revised Short Form**. *J ADV NURS* 2023, **79**(12):4732-4746.
22. Giammaria C, Panattoni N, Terrenato I, Spano A, De Leo A, Tomei B, Di Simone E, Petrone F: **Cross-cultural Adaptation and Validation of the Italian Version of the Revised Perceived Perioperative Competence Scale**. *J PERIANESTH NURS* 2025.
23. Yumei W: **Development of Competency Evaluation Index Development of Competency Evaluation Index System for Nurses in Operating Room**. □□. Huazhong University of Science and Technology, □□□; 2015.
24. Li M, Zheng W, Luping J, Qingxin F: **Development of the Operating Room Professional Nurses Core Competency Evaluation Scale**. *Chinese Journal of Modern Nursing* 2019, **25**(3):294-298.
25. Sevdalis N, Undre S, Henry J, Sydney E, Koutantji M, Darzi A, Vincent CA: **Development, initial reliability and validity testing of an observational tool for assessing technical skills of operating room nurses**. *INT J NURS STUD* 2009, **46**(9):1187-1193.
26. Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G: **Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system)**. *J EVAL CLIN PRACT* 2013, **19**(2):317-323.
27. yin S: **Reliability and validity of the Chinese Version of Scrub Practitioners' List of Intraoperative Non-Technical Skills**. □□. Tongji University, □□; 2020.
28. Sirevag I, Hansen BS, Tjoflat I, Gillespie BM: **Bi-lingual content validation of the Non-Technical Skills for Operating Room Nurses (NOTSORN) tool: A Delphi study**. *Int J Nurs Stud Adv* 2024, **7**:100218.
29. Kalantari R, Zamanian Z, Hasanshahi M, Jamali J, Faghihi AA, Niakan H, Abbasi J: **Development and psychometric evaluation of a behavioral marker system for circulating nurse's non-technical skills**. *Perioperative Care and Operating Room Management* 2021, **23**:100167.
30. Yu Q, Wei R, Wei Y, Wu X, Liang T: **Psychometric evaluation of the perceived perioperative competence scale-revised among the Chinese operating room nurses: a methodological research**. *BMC NURS* 2022, **21**(1):79.
31. Gillespie BM, Harbeck E, Sutherland Fraser S, Nicholson P, Boric T: **Psychometric validation of the Perceived Perioperative Competence Scale-Revised Short Form.** *Journal of Advanced Nursing (John Wiley & Sons, Inc.)* 2023, **79**(12):4732-4746.
32. Mundt AS, Spanager L, Lyk-Jensen HT, Østergaard D: **Adaptation of a tool to assess non-technical skills of scrub practitioners in Denmark**. *J Perioper Pract* 2017, **27**(9):180-185.
33. Mei N, Chang L, Zhu Z, Dong M, Zhang M, Zeng L: **Core competency scale for operating room nurses in China: Scale development, reliability and validity evaluation.** *NURS OPEN* 2022, **9**(6):2814-2825.
34. □□□: **CDIO**□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□. □□. □□□□□□□, □□; 2020.

35. Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G: **Evaluation of the Scrub Practitioners' List of Intraoperative Non-Technical Skills system**. *INT J NURS STUD* 2012, **49**(2):201-211.
36. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB: **COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study**. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018, **27**(5):1159-1170.
37. Liu M, Kunaiktikul W, Senaratana W, Tonmukayakul O, Eriksen L: **Development of competency inventory for registered nurses in the People's Republic of China: scale development**. *INT J NURS STUD* 2007, **44**(5):805-813.
38. Gillespie BM, Harbeck EB, Falk-Brynhildsen K, Nilsson U, Jaensson M: **Perceptions of perioperative nursing competence: a cross-country comparison**. *BMC NURS* 2018, **17**:12.

Figures

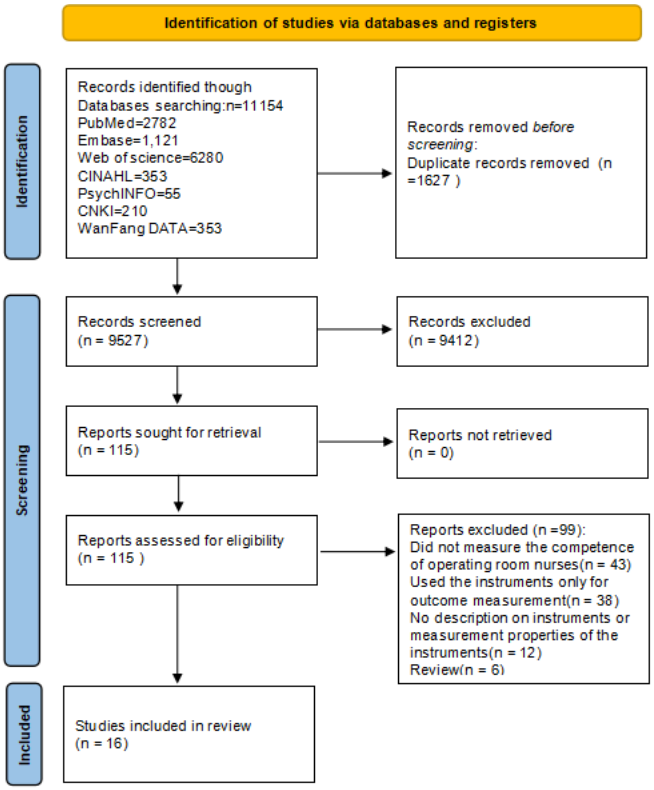


Figure 1

Flow diagram for systematic review.

Supplementary Files

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

- SupplementalAppendixASearchStrategyandSearchHistory.docx
- SupplementalAppendixBchecklist.docx