Table S1. Reasons for inconsistencies

Revised content	Reason
1. Review title	Taking into consideration the inclusive scope of the article, encompassing both nursing staff and nursing students, a deliberate revision has been undertaken in the title adjustment, transitioning from the initial term "nursing staff" to the more comprehensive descriptor "nursing personnel".
2. Funding	In the PROSPERO registration, the Funding sources/sponsors were initially documented as "None." However, it is important to note that during the course of the study, we secured pertinent funding support. Consequently, the Funding section in the protocol was subsequently revised to accurately reflect this development.

Table S2. Search strategy for PubMed

1	#1	nurs*
2	#2	((((((((((Health Education[Mesh] OR Patient Education as Topic[Mesh]) OR (health education*[tiab])) OR (education, health[tiab])) OR (patient education*[tiab])) OR (education program*[tiab])) OR (educational program*[tiab])) OR (education, patient*[tiab])) OR (education of patient*[tiab])) OR (nursing education*[tiab])) OR (hospital education*[tiab])) OR (educational activit*[tiab])) OR ((educat*[tiab]) AND (individual patient*[tiab]))
3	#3	((((((((professional competence[MeSH]) OR (professional competence[tiab])) OR (competence, professional[tiab])) OR (generalization of expertise[tiab])) OR (expertise generalization[tiab])) OR (technical expertise[tiab])) OR (expertise, technical[tiab])) OR (competenc*[tiab] OR capabilit*[tiab] OR capacit*[tiab] OR abilit*[tiab])) OR ((skill*[tiab] OR belief[tiab] OR attitude[tiab]) AND (knowledge[tiab]))
4	#4	(report[tiab] OR reported[tiab] OR reporting[tiab] OR rated[tiab] OR rating[tiab] OR ratings[tiab] OR based[tiab] OR assessed[tiab] OR assessment[tiab] OR assessments[tiab] OR disability[tiab] OR function[tiab] OR functional[tiab] OR functions[tiab] OR subjective[tiab] OR utility[tiab] OR utilities[tiab] OR wellbeing[tiab] OR well being[tiab]) AND (index[tiab] OR indices[tiab] OR instrument[tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR profile[tiab] OR profiles[tiab] OR scale[tiab] OR scales[tiab] OR score[tiab] OR scores[tiab] OR status[tiab] OR survey[tiab] OR surveys[tiab])

5	#5	(instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item

		functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])
6	#6	("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])
7	#7	#1 AND #2 AND #3 AND #4 AND #5 NOT #6

Note: "*" to include all derivatives of that word or concept.

Table S3. Search strategy for Chinese databases

1	#1	TKA = 护理 OR TKA = 护士 OR TKA = 护生
2	#2	TKA = 健康教育能力 OR TKA = 患者教育能力 OR TKA = 健康教育胜任力 OR TKA = 患者教育胜任力
3	#3	SU = 评估 OR SU = 测量 OR SU = 评价 OR SU = 收集 OR SU = 调查 OR SU = 工具 OR SU = 问卷 OR SU = 量表 OR SU = 仪器 OR SU = 研究
4	#4	TKA = 信度 OR TKA = 效度 OR TKA = 反应度 OR TKA = 内部一致性 OR TKA = 稳定性 OR TKA = 相关系数 OR TKA = 克朗巴赫系数 OR TKA = 探索性因子分析 OR TKA = 验证性因子分析 OR TKA = 探索性因素分析 OR TKA = 验证性因素分析 OR TKA = 检验 OR TKA = 结果
5	#5	#1 AND #2 AND #3 AND #4

Note: TKA = title/abstract; SU = title/abstract/keywords.

Table S4. Characteristics of the included instruments

Instrument name	Developer(s)/ year developed	Construct (s)	Target population	Mode of administration	Recall period	(Sub)scale (s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations

Table S5. Characteristics of the included study populations

		Population			Instrument administration			
Instrument	**Ref**	**N**	**Age Mean (SD, range) yr**	**Gender % female**	**Setting**	**Country**	**Language**	**Response rate**
A	1							
	2							
	3							
B	1							

Table S6. Rating the measurement properties of the instruments

Instrument	Study 1			Study 2			Study 3			OVERALL					
	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	OVERALL RATING	OVERALL RATING	OVERALL RATING	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE
	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	+ / - / ?	High, moderate, low, very low	High, moderate, low, very low	High, moderate, low, very low
	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus
Structural validity															
internal consistency															
Cross-cultural validity															
Measurement invariance															
Reliability															
Measurement error															
Criterion validity															
Construct validity															
Responsiveness															

Note: “+” = sufficient; “-” = insufficient; “?” = indeterminate.

Table S6-1. Rating of the content validity of instruments

Instrument (Reference – study type/Rating of reviewers)	**Content Validity**													
	Relevance[1]						**Comprehensiveness[1]**		**Comprehensibility[1]**					**CONTENT VALIDITY RATING[2]**
	1. Are the included items relevant for the construct of interest?	2. Are the included items relevant for the target population of interest?	3. Are the included items relevant for the context of use of interest?	4. Are the response options appropriate?	5. Is the recall period appropriate?	**RELEVANCE RATING[2]**	6. Are all key concepts included?	**COMPREHENSIVENESS RATING[2]**	7. Are the PROM instructions understood by the population of interest as intended?	8. Are the PROM items and response options understood by the population of interest as intended?	9. Are the PROM items appropriately worded?	10. Do the response options match the question?	**COMPREHENSIBILITY RATING[2]**	
A (Ref 1- instrument development study)														
A (Ref 2 - Content validity study)														
A (Ref 3 - Content validity study)														
Rating of reviewers														
B (Ref 1- instrument development study)														
B (Ref 2 - Content validity study)														
Rating of reviewers														

……														

Note:1. Rating for the 10 criteria for relevance, comprehensiveness, comprehensibility can be + / − /±/ ?: ‘ + ’= sufficient, ‘ − ’= insufficient, ‘±’ = inconsistent,‘?’ =indeterminate.

2. The RELEVANCE, COMPREHENSIVENESS, COMPREHESIBILITY, AND CONTENT VALIDITY rating can be + / − /±/ ? : ‘ + ’= sufficient, ‘ − ’= insufficient, ‘±’ = inconsistent, ‘?’ =indeterminate.

Table S7. Information on interpretability of instruments

Instrument (ref)		**Percentage of missing items and percentage of missing total scores**	**Floor and ceiling effects**	**Scores and change scores available for relevant (sub)groups**	**Minimal important change (MIC) or minimal important difference (MID)**	**Information on response shift**
Instrument A (ref 1)						
Instrument A (ref 2)						
Instrument A (ref 3)						
Instrument B (ref 1)						
……						

Table S8. Information on feasibility of instruments

Feasibility aspects	**Instrument A**	**Instrument B**	**Instrument C**	**Instrument D**
Patient's comprehensibility				
Clinician's comprehensibility				
Type and ease of administration				
Length of the instrument				
Completion time				
Patient's required mental and physical ability level				
Ease of standardization				
Ease of score calculation				
Copyright				
Cost of an instrument				
Required equipment				
Availability in different settings				
Regulatory agency's requirement for approval				

Table S9. Quality[1] of studies on measurement properties

Instrument	Content validity[2]					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Co
	Asking patients			Asking experts								
	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness							Converg validit
A												
B												
……												

Note: 1. Quality: V = very good, A = adequate, D = doubtful, I = inadequate.
2. Given that the criteria and rating systems for evaluating the content validity of instruments are different from those for other measurement properties, the quality results of content validity are not included in this table but separately shown in following Table S8-1.

Table S9-1. Quality[1] of the instrument development

Instrument	PROM design							Cognitive interview (CI) study[3]				TOTAL PROM DEVELOPMENT
	General design requirements					Concept elicitation[2]	Total PROM design	General design requirements	Comprehensibility	Comprehensiveness	Total CI study	
	Clear construct	Clear origin of construct	Clear target population for which the PROM was developed	Clear context of use	PROM developed in sample representing the target population			CI study performed in sample representing the target population				
A												
B												
……												

Note: 1. Quality: V = very good, A = adequate, D = doubtful, I = inadequate.
2.The concept elicitation will not be further rated if the instrument(s) was not developed in the sample representing the target population.
3. Empty cells indicate that a CI study (or part of it) was not performed.

Table S10. Quality of the evidence for measurement properties of the instruments (Summary of findings)

Instrument	Content validity		Structural validity		Internal consistency		Cross-cultural validity		Reliability		Measurement error		Criterion validity		Hypotheses testing		Responsiveness	
	Overall Rating[1]	Quality of Evidence[3]	Overall Rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]	Overall rating[2]	Quality of evidence[3]
Instrument A																		
Instrument B																		
Instrument C																		
……																		

Note:1. Overall ratings for the content validity (relevance, comprehensiveness, comprehensibility) can only be + / − /±: ‘ + ’= sufficient, ‘ − ’= insufficient, ‘±’ = inconsistent.

2. Overall ratings for other measurement properties can be + / − /±/ ?: ‘ + ’= sufficient, ‘ − ’= insufficient, ‘±’ = inconsistent, ‘?’ =indeterminate.

3. Ratings for quality of evidence: High, Moderate, Low, Very low.