Original Article

Access this article online
Quick Response Code:

Website: www.jehp.net
DOI: 10.4103/jehp.jehp_338_22

Development and assessment of validity and reliability of a checklist to evaluate the Circulating and Scrub Skills of Operating Room Novices (CSSORN checklist)

Morteza Nasiri, Shahrzad Yektatalab[1], Marzieh Momennasab[1], Fatemeh Vizeshfar[1]

Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran, [1]Community-Based Psychiatric Care Research Center, School of Nursing and Midwifery, Shiraz University of Medical Sciences, Shiraz, Iran

Address for correspondence:
Dr. Shahrzad Yektatalab, School of Nursing and Midwifery, Shiraz University of Medical Sciences, Shiraz, P.O. Box: 71936-13119, Iran.
E-mail: Yektash@sums.ac.ir

Received: 03-03-2022
Accepted: 14-06-2022
Published: 28-02-2023

Abstract:

BACKGROUND: Assessment of circulating and scrub skills is an important issue in operating room (OR) programs. However, there is a lack of well-designed tools that are specifically developed for this purpose. Hence, this study aimed to develop and determine the validity and reliability of a checklist to measure the circulating and scrub skills of OR novices.

MATERIALS AND METHODS: This cross-sectional methodological study was conducted among 124 OR technology students who were selected during three consecutive academic years (2019–2020 to 2021–2022). The developed checklist was evaluated with face validity, content validity (quantitative and qualitative), construct validity (known-groups validity), criterion-related validity (concurrent and predictive validities), internal consistency (Kuder–Richardson 20, KR-20), and inter-rater reliability (intra-class correlation coefficient, ICC). Known-groups validity was evaluated by comparing the difference between the checklist scores of first-semester and third-semester students using independent samples t-test. Additionally, concurrent and predictive validities were evaluated by ICC through measuring the correlation between the total score of checklist and grades of a multiple-choice test and two clinical apprenticeship courses, respectively. Data were analyzed in the Statistical Package for Social Sciences software.

RESULTS: After reconciling the preliminary checklist in terms of face and content validities, a checklist with 17 sub-scales and 340 items called "*Circulating and Scrub Skills of Operating Room Novices (CSSORN)"* was developed. Regarding the known-groups validity, the third-semester students had higher scores compared to the first-semester students ($p < 0.001$ in most sub-scales). Besides, the total score of checklist showed a significant correlation with the criteria of concurrent and predictive validities ($ICC = 0.64$, $ICC = 0.72$; $P < 0.001$). The KR-20 for the entire checklist was 0.90 (range: 0.60–0.93). The ICC for inter-rater reliability was also 0.96 for the entire checklist (range: 0.76–0.99, $P < 0.001$ in all sub-scales).

CONCLUSION: The *CSSORN* had appropriate validity and reliability to be used for measuring the circulating and scrub skills of OR novices. To shed light on the findings, further testing of this checklist on larger populations and in different contexts is suggested.

Keywords:

Checklist, clinical competence, operating room nursing, psychometrics, students

Introduction

Assessment of clinical skills is one of the most difficult responsibilities of instructors and a challenge for educational programs.[1] The importance of clinical skills assessment is more critical when students are prepared to provide health care in

How to cite this article: Nasiri M, Yektatalab S, Momennasab M, Vizeshfar F. Development and assessment of validity and reliability of a checklist to evaluate the Circulating and Scrub Skills of Operating Room Novices (CSSORN checklist). J Edu Health Promot 2023;12:65.

high-risk and complex environments such as operating rooms (ORs).[2] Crucially, students need to show that they can perform the peri-operative procedural tasks, which is achievable as the best by observation of their performance that conforms to the "shows how" and "does" levels of Miller's pyramid of assessment.[3]

A proper observational tool can increase the likelihood of performing a systematic and valid assessment in OR fields and makes a noticeable contribution to the learning process through constructive feedback on students' performance and ensuring the instructors about whether or not the students have obtained the required skills to work in an OR.[4,5] However, there is no definite tool to assess the students' skills in OR fields because the development of tools depends on the context and job description of the OR team members; thus, a variety of indirect and direct measures are applied in many countries including Iran to address this issue.[6-9]

Despite most countries with OR nursing as a branch of the nursing specialty, in Iran, a 4-year centralized program called *"OR technology"* is currently presented as a separate branch of medical education.[10] This program aims to prepare students for acting as OR technologists in specialized and sub-specialized surgeries in two roles of a circulating person and scrub person.[11] Given the ever-increasing society's expectations of quality care, the focus of *"OR technology"* education has been on students' procedural skills and competencies.[12] However, the assessment of peri-operative procedural skills has been criticized in Iran as it is performed commonly arbitrary and subjective, and no accurate, objective, and clear criteria are used for this purpose.[13] Therefore, developing valid and reliable instruments to determine procedural or psychomotor skills of Iranian OR technology students is needed and deserves further study.

Previous studies have developed different observational or self-report tools to measure the procedural, technical, and non-technical skills of Iranian students in OR fields.[12,14,15] However, to the best of our knowledge, a standardized tool has not been established to specifically measure students' circulating and scrub skills, which are fundamental for working as an OR technologist. Thus, the aim of this study is to develop and assess the validity and reliability of a checklist in Persian to assess the circulating and scrub skills of first-semester OR technology students before they begin their apprenticeship training courses.

Materials and Methods

Study design and setting

This cross-sectional methodological study was conducted in three phases, development, validity evaluation (i.e., face, content, construct, and criterion-related), and reliability estimation (i.e., inter-rater reliability, IRR; and internal consistency) [Figure 1].

Study participants and sampling

To evaluate the IRR of the checklist with the same raters and similar conditions, the statistical population of the study was selected only from Shiraz University of Medical Sciences (SUMS), Shiraz, Iran. To reduce bias by increasing the sample size, the study was conducted over three consecutive academic years from 2019–2020 to 2021–2022. A total of 124 OR technology students were selected via the census method (91 first-semester and 33 third-semester students).

Phase I: Development

Selecting circulating and scrub skills

To select circulating and scrub skills, 20 skills that a first-semester OR technology student must acquire in the role of a circulating or scrub person were listed based on the Iranian undergraduate OR curriculum for the course titled *"Principles and Techniques of Circulating and*

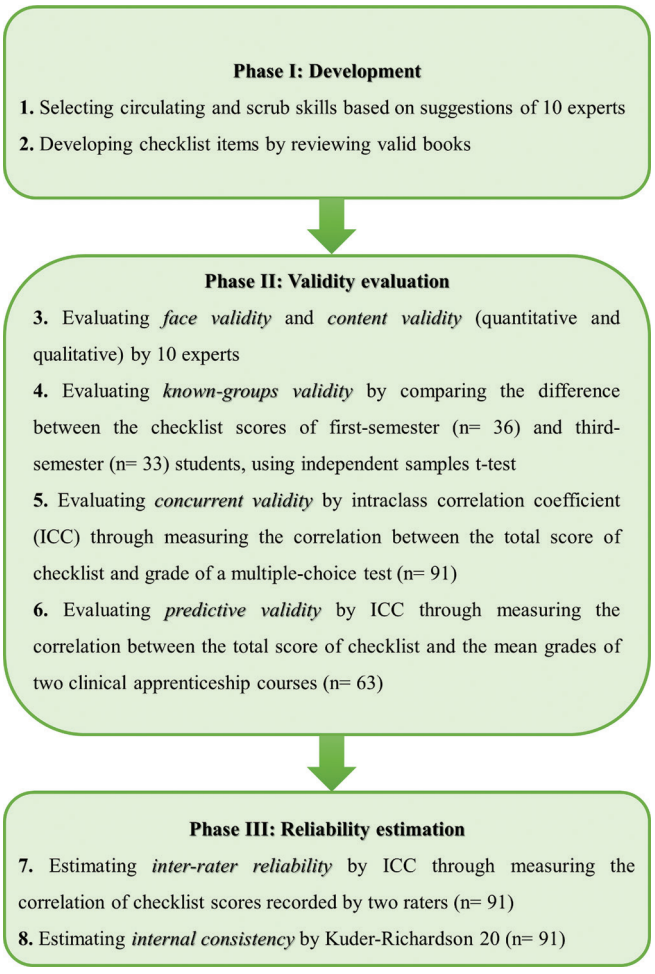

Figure 1: The process of development and assessment of validity and reliability of the *CSSORN* checklist designed to evaluate the circulating and scrub skills of OR novices

Scrub Persons' Practice".[11] Then, to prioritize the skills, the prepared list was presented to ten experts who were well experienced in terms of the study objectives and had at least 5 years of working experience. The experts were selected from Iranian universities of medical sciences using purposive sampling from August to December 2019. The experts were requested to rank the listed skills based on their importance. Out of 20 listed skills, 17 high-rank skills were selected based on the experts' suggestions. Also, each skill was considered as a checklist sub-scale based on the consensus achieved among the experts.

Developing checklist items

To develop the checklist items, the actions expected from students to perform for each selected skill were written step by step based on the books on fundamentals of circulating and scrub skills, suggested by the Iranian undergraduate OR curriculum.[16-18] Then, each step was considered as a checklist item. A total of 332 items were extracted for 17 selected skills. Accordingly, the preliminary checklist with 17 sub-scales and 332 items was developed. The number of items varied depending on the sub-scales. Based on the experts' suggestions, the response scale for each item was built with a binary option: 1) performed: score 1 and 2) not performed: score 0.

Phase II: Validity evaluation

Evaluating face and content validities

To measure the face validity and content validity (qualitative and quantitative), the preliminary checklist was sent to the ten experts who prioritized the 20 circulating and scrub skills in the first stage. The experts were asked to investigate the checklist items qualitatively in terms of face validity criteria (i.e., difficulty, ambiguity, and syntax) as well as content validity criteria (i.e., objectivity, number of items, the logical sequence of items, and scoring).[19] Most experts were satisfied with the qualitative face validity and qualitative content validity of the preliminary checklist. However, some experts suggested extra items or minor changes, which resulted in modifying the preliminary checklist in two sequential steps. After reaching a consensus among the experts, the second version of the checklist was established, which consisted of 343 items.

To evaluate the quantitative content validity of the second version of the checklist, the content validity ratio (CVR) and content validity index (CVI) were determined. To estimate the necessity of each item, the CVR values were calculated by the method proposed by Lawshe.[20] To this end, the experts were asked to score each item on a 3-point scale (1: unnecessary, 2: useful but unnecessary, and 3: necessary). Then, the CVR was calculated by the following equation: CVR = [ne–(N/2)]/(N/2), in which the *"ne"* is the number of experts who identified the item as "necessary" and *"N"* is the total number of experts. Next, items with a CVR higher than 0.62 were retained. To investigate the relevance of each item, the CVI values were calculated by Waltz and Bausell's approach.[21] For this purpose, the experts were asked to rank each item in accordance with its relevancy by using a 4-point scale (1: not relevant, 2: somewhat relevant, 3: quite relevant, and 4: very relevant). Then, item-level CVI (I-CVI) was calculated by the number of experts who identified the item as either "very relevant" or "quite relevant" divided by the total number of experts.[22] If the score of I-CVI was more than or equal to 0.79, the item was retained in the checklist.[23] To obtain scale-level CVI (S-CVI), the average scale CVI (S-CVI/Ave) was estimated by averaging the I-CVI values. An S-CVI/Ave of 0.90 or higher is considered evidence of "excellent" content validity.[22]

Evaluating construct validity (known-groups approach)

Construct validity is usually determined using the hypothesis testing approach, the multi-trait–multi-method approach, the known-groups approach, and/or factor analysis.[19] Out of these, the known-groups approach, which has also been called discriminative validity and contrast validity, is commonly used when the instrument's response scale is categorical with less than five options.[24] This approach relies on hypotheses concerning an instrument's ability to discriminate between two or more groups known (or expected) to differ in the construct of interest.[25] Based on this approach, investigating whether there are differences between the groups can provide evidence of the construct validity of the assessment instrument.[26]

To measure known-groups validity in the current study, the checklist scores of first-semester and third-semester under-graduate OR technology students were compared because it is expected that more experience would result in better performance of peri-operative skills.[27] To this end, 36 first-semester students taking the training course titled *"Principles and Techniques of Circulating and Scrub Persons' Practice"* and 33 third-semester students taking the clinical apprenticeship course titled *"Principles and Techniques of Circulating Person"* in the academic year 2019–2020 participated in a practical examination, and their circulating and scrub skills were evaluated and scored through direct observation using the developed checklist by two trained raters independently. The mean scores recorded by the two raters were used to estimate the known-groups validity.

Evaluating concurrent validity

To determine the criterion-related validity by the concurrent approach, circulating and scrub skills of 91 first-semester under-graduate OR technology

students were evaluated using the developed checklist and a multiple-choice question (MCQ) test; then, the correlation of the total score of the checklist with the grade of MCQ test was estimated. The MCQ test included 20 questions and was held as a routine method for summative evaluation of the practical credit of the course titled *"Principles and Techniques of Circulating and Scrub Persons' Practice"*. The checklist was also completed through direct observation by two trained raters independently, and the mean scores recorded by the raters were used to estimate the concurrent validity. The students in this stage were selected during academic years 2019–2020 (n = 36), 2020–2021 (n = 27), and 2021–2022 (n = 28). To increase the sample size, 36 students whose data were recorded for estimation of known-groups validity during the academic year 2019–2020 were also included.

Evaluating predictive validity

To measure the predictive validity, the students were followed until their fourth semester and their grades in two clinical apprenticeship courses titled *"Principles and Techniques of Circulating Person"* and *"Principles and Techniques of Scrub Person"* were recorded in their third and fourth semesters, respectively. Then, the mean grades of the courses was computed, and their correlation with the total score of the checklist was estimated. Out of 91 students whose data were recorded for estimation of the concurrent validity, 63 students were included in this stage during the academic years 2020–2021 (n = 36) and 2021–2022 (n = 27). Other 28 students, whose data were recorded for estimation of concurrent validity during the academic year 2021–2022 could not be included in this stage because they will enter their third and fourth semesters in the academic year 2022–2023.

Phase III: Reliability estimation

The study population in this phase consisted of 91 students whose data were recorded for estimation of concurrent validity.

Estimating inter-rater reliability

To evaluate the IRR of the checklist, the circulating and scrub skills of students were scored using the developed checklist by two trained raters independently; then the correlation of checklist scores recorded by the two raters was estimated. The raters were the same for 91 students during three academic years to limit any bias.

Estimating internal consistency

To measure the internal consistency, use was made of the mean scores recorded by the two raters.

Data analysis

All data were analyzed using the Statistical Package for Social Sciences software (SPSS, version 22.00; SPSS Inc., USA). The normal distribution of data was confirmed using the Kolmogorov–Smirnov test. Accordingly, the independent samples t-test and Chi-square test were used to compare data of first-semester and third-semester students in the know-groups validity stage. To obtain the criterion-related validity and IRR results, the intra-class correlation (ICC) with consistency type and absolute agreement type were used, respectively. The following classification was used to interpret the ICC values: weak reliability, ICC < 0.50; moderate reliability, ICC 0.50–0.75; good reliability, ICC 0.75–0.90; and excellent reliability, ICC ≥ 0.90.[28] Given that the checklist was scored as performed/not performed, Kuder–Richardson 20 (KR-20) was estimated to measure the internal consistency, and a value of 0.70 or more was considered satisfactory.[29] A P *value* < 0.050 was considered as significant.

Ethical considerations

Ethical approval was obtained from the Local Research Ethics Committee of SUMS (Approval No. IR.SUMS.REC.1399.580). All the experts and students were informed about the study objectives and procedures, and their written informed consent forms were obtained. Moreover, the students were assured that their scores in the developed checklist did not constitute a part of their academic evaluation and the data were only used for study purposes. Finally, they were informed about their scores and also provided with incentives.

Results

Description of the participatns

The mean age of the 36 first-semester and 33 third-semester students who participated in the known-groups validity stage was 20.22 ± 1.65 and 20.93 ± 1.11 years, respectively (t = -2.08, P = 0.041). Regarding gender, 55.6% of the first-semester students were male, whereas 51.5% of third-semester students were female (χ^2 = 0.34, P = 0.557). On the other hand, 91 first-semester students who participated in the concurrent validity and reliability evaluation consisted of 48 males (52.7%) and 43 females (47.3%) with a mean age of 20.25 ± 1.72 years. Similarly, 63 first-semester students who participated in the predictive validity stage consisted of 35 males (55.6%) and 28 females (44.4%) with a mean age of 20.14 ± 1.38 years.

Quantitative content validity

The experts (n = 10) determined 340 items of the second version of the checklist as essential (*CVR:* 0.80–1.00). However, three items with a CVR lower than 0.62 were removed. Based on the results of I-CVI, all items were scored from 0.80 to 1.00. The S-CVI/Ave also ranged from

0.97 to 1.00. Accordingly, the last version of the checklist with 17 sub-scales and 340 items called *"Circulating and Scrub Skills of Operating Room Novices (CSSORN)"* was developed [Tables 1 and 2].

Checklist scoring

The checklist completion lasted about 30 min for students who participated in the practical examination for estimation of know-groups validity, criterion-related validity, and reliability. To evaluate the students' performance in circulating and scrub skills, each checklist item is scored on a binary format as "performed" (score 1) and "not performed" (score 0). Then, the total score of each sub-scale is computed by summing up the scores assigned to the items that make up that sub-scale. Subsequently, the total score of the checklist is calculated by summing up scores of 17 sub-scales, ranging from 0 to 340. Finally, the raw scores are all divided into 17 to calculate the total score in the range of 0–20 based on the Iranian under-graduate OR curriculum. The higher scores indicate a better level of circulating and scrub skills [Table 1].

Table 1: Content validity and reliability properties of the *CSSORN* checklist developed to evaluate the circulating and scrub skills of OR novices

Sub-scales	Number of items (score)	Content validity (*n*=10)			Reliability (*n*=91)		
		CVR	I-CVI	S-CVI/Ave	Internal consistency	Inter-rate reliability	
					KR-20	ICC (95% CI)	*P*
Donning surgical cap*	11 (0-0.65)	0.80-1.00	0.80-1.00	0.98	0.70	0.88 (0.78-0.94)	<0.001
Donning surgical face mask	14 (0-0.82)	0.80-1.00	0.90-1.00	0.98	0.71	0.79 (0.63-0.89)	<0.001
Scrubbing hands with povidone-iodine solution	51 (0-3.00)	1.00	0.90-1.00	0.99	0.80	0.90 (0.81-0.94)	<0.001
Drying scrubbed hands with the sterile towel	22 (0-1.30)	1.00	1.00	1.00	0.91	0.88 (0.63-0.89)	<0.001
Doffing a surgical cap and face mask	12 (0-0.70)	0.80-1.00	0.80-1.00	0.97	0.79	0.85 (0.73-0.92)	<0.001
Adding surgical gloves to an established sterile field	14 (0-0.82)	0.80-1.00	1.00	1.00	0.71	0.88 (0.77-0.93)	<0.001
Donning a surgical gown	24 (0-1.41)	0.80-1.00	1.00	1.00	0.75	0.90 (0.82-0.95)	<0.001
Donning surgical gloves by a close technique	27 (0-1.59)	1.00	1.00	1.00	0.76	0.84 (0.70-0.92)	<0.001
Doffing a surgical gown and gloves	17 (0-1.00)	1.00	1.00	1.00	0.73	0.76 (0.53-0.87)	<0.001
Placing an electrocautery safety plate	13 (0-0.77)	1.00	1.00	1.00	0.93	0.90 (0.82-0.95)	<0.001
Placing surgical drapes	34 (0-2.00)	1.00	1.00	1.00	0.75	0.90 (0.81-0.94)	<0.001
Removing of surgical drapes and the electrocautery safety plate	14 (0-0.82)	1.00	1.00	1.00	0.65	0.90 (0.83-0.95)	<0.001
Opening a pack of sterile surgical instrument tray sets	20 (0-1.18)	1.00	1.00	1.00	0.70	0.88 (0.79-0.94)	<0.001
Counting X-ray detectable sponges	15 (0-0.89)	1.00	1.00	1.00	0.78	0.99 (0.98-0.99)	<0.001
Adding the solution to an established sterile field	12 (0-0.70)	1.00	1.00	1.00	0.60	0.94 (0.87-0.97)	<0.001
Prepping the surgical site skin	18 (0-1.05)	1.00	1.00	1.00	0.63	0.97 (0.95-0.98)	<0.001
Wrapping the surgical instrument tray set	22 (0-1.30)	1.00	1.00	1.00	0.76	0.96 (0.92-0.98)	<0.001
Total	340 (0-20.00)	-	-	-	0.90	0.96 (0.93-0.98)	<0.001

CI, confidence interval; CVR: content validity ratio; ICC, intra-class correlation coefficient; I-CVI: item-level content validity index; KR-20: Kuder-Richardson 20 coefficient; S-CVI/Ave: scale-level content validity index/average* Content validity of items developed for this sub-scale has been presented in Table 2 with details

Table 2: Content validity ratio and content validity index of items developed for the "*donning surgical cap*" sub-scale of the *CSSORN* checklist

Items	CVR	I-CVI
Examines the surgical cap for appearance integrity (i.e., moisture, tearing, contamination) through observation and touching.	0.90	1.00
Replaces the surgical cap correctly if there is a problem with its appearance integrity (if there is no problem, states this orally).	0.80	1.00
Puts the surgical cap on her/his head so that the cap straps are placed behind her/his head.	1.00	1.00
While the surgical cap is put on her/his head, pulls it down to cover her/his hair.	1.00	1.00
Folds the edges of the surgical cap gently, if necessary (i.e., looseness). If there is no need, states this orally.	0.80	1.00
Wears the surgical cap so that maximum coverage is created on her/his hair.	1.00	1.00
Ties the surgical cap straps at the back of her/his head so that it is not too tight or loose.	1.00	1.00
Checks her/his appearance in the mirror after donning the surgical cap to make sure it completely covers her/his hair.	1.00	1.00
Moves the part of her/his hair that is left out of the surgical cap under the cap by hand.	1.00	1.00
Wears the surgical cap so that its outer surface is outside and its inner surface is inside (not upside down).	1.00	1.00
Performs the given task at the appointed time (maximum 45 sec)	1.00	0.80

CVR: content validity ratio; I-CVI: item-level content validity index

Known-groups validity

Known-groups validity was assessed by comparing the checklist scores between 36 first-semester and 33 third-semester students. The total score of the checklist was significantly higher in the third-semester students compared to the first-semester students ($p < 0.001$). Also, the obtained scores in all sub-scales were significantly higher for the third-semester students compared to the first-semester students ($p < 0.001$ in most cases) [Table 3].

Concurrent validity

To measure the concurrent validity of the checklist among 91 first-semester students, the MCQ test was used as a reference (score: 0–20). The qualitative content validity of the MCQ test was confirmed by ten experts who participated in evaluating the checklist's face and content validation. Also, the reliability of this test was approved among 30 first-semester OR technology students (not included in the main analysis) through internal consistency (*KR-20* = 0.73) and 14-day test–re-test reliability (*ICC* = 0.95, *95% confidence interval*: 0.86–0.94, $P < 0.001$). The mean grade of the MCQ test and the total score of the checklist among 91 students were 15.29 ± 2.27 and 17.38 ± 1.38, respectively. A significant correlation was observed between the total checklist score and grade of the MCQ test (*ICC* = 0.64, *95% confidence interval*: 0.45–0.76, $P < 0.001$).

Predictive validity

To measure the predictive validity of the checklist among 63 first-semester students, two clinical apprenticeship courses were used as a reference (grade: 0-20). The mean grade of the two clinical apprenticeship courses and the total score of the checklist were 15.30 ± 3.27 and 17.26 ± 1.61, respectively. A significant correlation was found between the total checklist score and the mean grades of the two clinical apprenticeship courses (*ICC* = 0.72, *95% confidence interval*: 0.54–0.83, $P < 0.001$).

Reliability

The reliability of the checklist was estimated among 91 first-semester students. The ICC coefficient for IRR was 0.96 for the entire checklist and ranged from 0.76 to 0.99 for the sub-scales ($p < 0.001$ in all cases). Likewise, the KR-20 coefficient was 0.90 for the entire checklist and ranged from 0.60 to 0.93 for the sub-scales [Table 1].

Discussion

The OR novices need to obtain the required circulating and scrub skills to provide safe patient care before, during, and after surgical procedures.[30] A review of the literature indicates a lack of instruments that are specifically developed to evaluate the circulating and scrub skills. Thus, a checklist was developed to cover this gap, and subsequently, its properties were evaluated by measuring the face, content, construct, and criterion-related validities as well as IRR and internal consistency.

To establish the content validity of the checklist, the steps of 17 important skills that an OR technology

Table 3: Comparison of mean (standard deviation) of circulating and scrub skills of first-semester and third-semester OR technology students who participated in the known-groups validity evaluation of the *CSSORN* checklist

Sub-scales	Mean±SD		*t*	*P*
	First-semester students (*n*=36)	Third-semester students (*n*=33)		
Donning surgical cap (score: 0-0.65)	0.50±0.10	0.59±0.05	- 4.16	<0.001
Donning surgical face mask (score: 0-0.85)	0.62±0.10	0.79±0.04	- 8.68	<0.001
Scrubbing hands with povidone-iodine solution (score: 0-3.00)	2.71±0.23	2.92±0.08	- 4.77	<0.001
Drying scrubbed hands with the sterile towel (score: 0-1.30)	0.75±0.07	1.23±0.14	- 17.65	<0.001
Doffing a surgical cap and face mask (score: 0-0.70)	0.51±0.11	0.60±0.07	- 4.18	<0.001
Adding surgical gloves to an established sterile field (score: 0-0.82)	0.56±0.13	0.78±0.05	- 8.93	<0.001
Donning a surgical gown (score: 0-1.41)	1.16±0.13	1.34±0.10	- 6.00	<0.001
Donning surgical gloves by a close technique (score: 0-1.59)	1.32±0.11	1.56±0.07	- 9.87	<0.001
Doffing a surgical gown and gloves (score: 0-1.00)	0.83±0.12	0.99±0.01	- 6.88	<0.001
Placing an electrocautery safety plate (score: 0-0.77)	0.44±0.13	0.62±0.09	- 6.25	<0.001
Placing surgical drapes (score: 0-2.00)	1.58±0.15	1.89±0.11	- 9.41	<0.001
Removing of surgical drapes and the electrocautery safety plate (score: 0-0.82)	0.62±0.10	0.76±0.09	- 5.71	<0.001
Opening a pack of sterile surgical instrument tray sets (score: 0-1.18)	1.04±0.13	1.12±0.05	- 2.89	0.005
Counting X-ray detectable sponges (score: 0-0.89)	0.69±0.15	0.80±0.10	- 3.13	0.003
Adding the solution to an established sterile field (score: 0-0.70)	0.53±0.12	0.60±0.07	- 2.61	0.011
Prepping the surgical site skin (score: 0-1.05)	0.91±0.12	0.97±0.08	- 2.31	0.023
Wrapping the surgical instrument tray set (score: 0-1.30)	1.15±0.14	1.22±0.07	- 2.13	0.037
Total (score: 0-20.00)	16.01±0.86	18.75±0.99	- 12.24	<0.001

student must perform during his/her first semester were determined by reviewing the valid books. Also, the face and content validities of the preliminary checklist were assessed by a multi-disciplinary team (i.e., six instructors of OR, two medical education professionals, and two psychometricians). Out of 343 primary items, three items with a CVR lower than 0.62 were deleted. Of the remaining 340 necessary items, all had good relevancy based on the I-CVI values. Also, the S-CVI/Ave of the checklist was excellent, indicating high inter-rater agreement over the relevance of the items. Accordingly, the final version of the checklist with 340 items and 17 sub-scales called the *CSSORN* was examined for construct and criterion-related validities as well as reliability.

The results of known-groups validity indicated a significant difference in the obtained scores between the two groups of students; the first-semester students had lower scores than the third-semester students, which signifies good construct validity of the *CSSORN* checklist. Also, the results of concurrent validity showed a moderate correlation between the total score of the *CSSORN* checklist and the grade of the MCQ test held for summative evaluation of the circulating and scrub skills. Hence, the *CSSORN* checklist could be an acceptable alternative to the traditional MCQ test held commonly for evaluating the circulating and scrub skills. Similarly, based on the results of predictive validity, a good correlation was found between the total score of the *CSSORN* checklist and the mean grades of two courses held for the apprenticeship of the circulating and scrub skills. It means that the *CSSORN* checklist had a good ability to predict the future performance of OR technology students in circulating and scrub tasks. Additionally, the results of IRR revealed good-to-excellent agreement between the scores recorded by the two raters. The highest IRR was related to the *"counting X-ray detectable sponges"* sub-scale, whereas the lowest IRR was related to the *"doffing a surgical gown and gloves"* sub-scale. Likewise, the results revealed an acceptable level of internal consistency for the entire checklist, indicating that all items were related and measured the same construct. Out of 17 sub-scales, 14 sub-scales had an internal consistency above the satisfactory level ranging from 0.70 to 0.93. The remaining three sub-scales had a moderate internal consistency in the range of 0.60–0.65, which might be because of the small sample size or the low number of items within the sub-scale (*"adding the solution to an established sterile field"*: 12 items, *"prepping the surgical site skin"*: 18 items, and *"removing of surgical drapes and the electrocautery safety plate"*: 14 items).

The findings of the present study corroborated those of a few studies documenting the validity and reliability of tools to evaluate peri-operative skills. Although previous studies have developed instruments to measure the circulating and scrub skills, they focus on general peri-operative skills and do not cover all aspects of the circulating and scrub skills.[14,15,31] Mahdavi *et al*.[15] developed a 38-item observational tool to measure the clinical skills of nursing students in OR, which included four sub-scales of *"responsibilities of scrub nurse"* (14 items), *"responsibilities of circulating nurse"* (nine items), *"recovery room care of patients"* (eight items), and *"sterilization and infection control"* (seven items). Similar to the present study, they obtained a minimum CVR and CVI of 0.80 for each item and an internal consistency of 0.85 for the entire questionnaire among 45 nursing students. However, compared to the present study, they did not measure the S-CVI/Ave, the criterion-related and construct validities, and the IRR. In another study, Mohammadi *et al*.[14] developed a 62-item observational tool for the clinical evaluation of OR students. Out of seven sub-scales, five items of the *"clinical competence"* sub-scale were related to the circulating and scrub skills (i.e., surgical scrub, surgical growing and gloving, surgical prep, surgical drape, and surgical pack opening). They obtained a CVR and I-CVI of 0.69–0.88 and an S-CVI of 0.84. With a sample of 28 students, they found an internal consistency of 0.82 for the entire tool, whereas with a sample of 15 students, the IRR was 0.45 for the entire tool. In the current study, the sample size was higher in the reliability stages compared to the aforesaid study. Also, the differences might be because of discrepancies in students' knowledge, experience, skill, culture, and differences in curriculum emphasis between the two target populations and also the tool's objectives. Compared to the present study, the criterion-related and construct validities were not evaluated in the aforesaid study.

Study implications

The *CSSORN* checklist could be used as a new observational tool in education and research to evaluate the circulating and scrub skills of OR novices through direct observation. It is of practical use before the beginning of the internship or at the end of the internship to evaluate the students' competency in performing the circulating and scrub tasks. The *CSSORN* checklist will help the instructors to quantify what steps of a circulating and scrub task need more instruction. Also, the checklist can familiarize the novices with the order of the steps in a determined task and might encourage them to improve their skills by performing the steps in order.

Limitation and recommendation

The results of this study may not be representative of the general population because of the sampling in a single university in Iran; thus, further studies in other Iranian universities are recommended. Furthermore, it was developed based on the Iranian under-graduate OR curriculum and Iranian context. Maybe in some other

countries, the job description of the circulating person and scrub person is different, which limits the generalizability of the checklist. However, as the items were derived from international valid books, which are similar throughout the world, it seems likely that the *CSSORN* checklist is generalizable to other countries. Accordingly, more studies are suggested on other cultures, languages, and contexts for cultural adaptation and more accurate evaluations of the reliability and validity of the checklist. Additionally, the high number of checklist items might make it difficult to use. The best option is to use all sub-scales together; however, it seems that each sub-scale could be used alone to measure a specific skill. Thus, further studies are recommended to measure the validity and reliability indices of each sub-scale to better understand the usability of the sub-scales. Finally, we did not evaluate convergent validity because of the lack of a well-established measure of circulating and scrub skills; thus, further study in this regard is suggested.

Study novelty

To the best of the authors' knowledge, this is the first attempt in Iran at developing and evaluating validity and reliability of a specific observational tool in terms of circulating and scrub skills. The most important strength of the developed checklist was to comprise the most fundamental circulating and scrub tasks, which help to a comprehensive evaluation of OR novices' circulating and scrub skills. Moreover, in this study, an adequate sample size and the application of different approaches to validity and reliability testing, including content validity, construct validity, criterion-related validity, and reliability, contributed to the development of an adequately valid checklist for the measurement of circulating and scrub skills.

Conclusions

Optimizing circulating and scrub skills can enhance the quality and effectiveness of peri-operative care, thereby increasing the satisfaction of the providers and receivers of health care. Accordingly, it is necessary to use a valid tool to evaluate the required OR novices' circulating and scrub skills. The results from this preliminary study suggest good validity and reliability of the *CSSORN* checklist for the assessment of the circulating and scrub skills of the first-semester OR technology students. Thus, using this checklist, OR instructors and officials can identify the weakness in circulating and scrub skills of OR novices, and by modifying them using appropriate measures, the quality of education and also patient care during the peri-operative period can be promoted.

Acknowledgements

The present article was extracted from a thesis written by Morteza Nasiri for obtaining the Ph.D. degree in Nursing at Shiraz University of Medical Sciences, Shiraz, Iran (approval No. 98-01-08-21354). The authors wish to express their gratitude to the experts who cooperated in evaluating the face and content validities of the checklist. They would also like to thank all the students who took part in the construct validity, concurrent validity, and reliability stages.

Financial support and sponsorship

This study was financially supported by the Vice-Chancellor of Research of Shiraz University of Medical Sciences, Shiraz, Iran (approval No. 98-01-08-21354).

Conflicts of interest

There are no conflicts of interest.

References

1. Christiansen B, Averlid G, Baluyot C, Blomberg K, Eikeland A, Strand Finstad IR, *et al*. Challenges in the assessment of nursing students in clinical placements: Exploring perceptions among nurse mentors. Nurs Open 2021;8:1069-76.
2. Heidarpoor P, Yazdani S, Sadati L. Non-technical skill assessment tools for surgical team members in the operating room: A critical review. Shiraz E-Med J 2021;22:e101291.
3. Ten Cate O, Carraccio C, Damodaran A, Gofton W, Hamstra SJ, Hart DE, *et al*. Entrustment Decision Making: Extending Miller's Pyramid. Acad Med 2021;96:199-204.
4. Bui AH, Guerrier S, Feldman DL, Kischak P, Mudiraj S, Somerville D, *et al*. Is video observation as effective as live observation in improving teamwork in the operating room? Surgery 2018;163:1191-6.
5. Schreyer J, Koch A, Herlemann A, Becker A, Schlenker B, Catchpole K, *et al*. RAS-NOTECHS: Validity and reliability of a tool for measuring non-technical skills in robotic-assisted surgery settings. Surg Endosc 2022;36:1916-26.
6. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current status of technical skills assessment tools in surgery: A systematic review. J Surg Res 2020;246:342-78.
7. Fahim C, Wagner N, Nousiainen MT, Sonnadara R. Assessment of technical skills competence in the operating room: A systematic and scoping review. Acad Med 2018;93:794-808.
8. Hannani S, Arbkhazaie A, Sadati L, Arbkhazaie A. Determining operating room technologists' clinical competence in educational care hospitals of Iran University of Medical Sciences in the year is 2017. J Adv Pharm Edu Res 2019;9:70-5.
9. Kalantari R, Zamanian Z, Hasanshahi M, Jamali J, Faghihi AA, Niakan H, *et al*. Development and psychometric evaluation of a behavioral marker system for circulating nurse's non-technical skills. Perioper Care Oper Room Manag 2021;23:100167.
10. Zardosht R, Karimi Moonaghi H, Etezad Razavi M, Ahmady S. Educational concern of surgical technology students in the operating room: A grounded theory study. J Educ Health Promot 2020;9:58.
11. SCPMS. B.Sc. of Operating Room Technology Curriculum. Tehran, Iran: Ministery of Health and Medical Education of Iran; 2014.
12. Mirbagher Ajorpaz N, 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;25:162-72.
13. Yazdimoghaddam H, Samadipour E, Ghardashi F, Borzoee F, Akbarzadeh R, Zardosht R, *et al*. Designing a comprehensive clinical competency test for operating room technology student: Using Delphi technique and CIPP model evaluation. J Educ Health

Promot 2021;10:240.
14. Mohammadi S, Tolyat M, Roshanzadeh M. Clinical evaluation tool for operating room students: Development and measurement of reliability and validity. Iran J Med Educ 2015;15:98-110.
15. Mahdavi R, Pooladi S, Bahreini M, Motamed N, Hajinezhad F. Designing and determining the psychometric properties of the clinical skills questionnaire for operating room nursing students. Strides Dev Med Educ 2017;14:e68126.
16. AST. Surgical Technology for the Surgical Technologist: A Positive Care Approach. Fifth ed. Cengage Learning; 2017.
17. Phillips NF. Berry and Kohn's Operating Room Technique. Thirteen ed. Elsevier Mosby; 2017.
18. Rothrock JC. Alexander's Care of the Patient in Surgery. Sixteenth ed. Elsevier Health Sciences; 2018.
19. Waltz C, Strickland OL, Lenz E. Measurement in Nursing and Health Research. 5th ed. New York: Springer Publishing Company; 2016.
20. Lawashe CH. A quantitative approach to content validity. Pers Psychol 1975;28:563-75.
21. Waltz CF, Bausell RB. Nursing Research: Design, Statistics, and Computer Analysis. Philadelphia: F.A. Davis Company; 1983.
22. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007;30:459-67.
23. Torabizadeh C, Yousefinya A, Zand F, Rakhshan M, Fararooei M. A nurses' alarm fatigue questionnaire: Development and psychometric properties. J Clin Monit Comput 2017;31:1305-12.
24. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. Eight ed. Pearson Education Limited; 2018.
25. Polit DF, Yang F. Measurement and the Measurement of Change. Philadelphia, PA: Wolters Kluwer; 2016.
26. Kasujja R, Bangirana P, Chiumento A, Hasan T, Jansen S, Kagabo DM, *et al*. Translating, contextually adapting, and pilot testing of psychosocial and mental health assessment instruments for Congolese refugees in Rwanda and Uganda. Confl Health 2022;16:17.
27. Nasiri M, Eslami J, Rashidi N, Paim CPP, Akbari F, Torabizadeh C, *et al*. "Playing with Surgical Instruments (PlaSurIn)" game to train operating room novices how to set up basic surgical instruments: A validation study. Nurse Educ Today 2021;105:105047.
28. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for Reliability Research. J Chiropr Med 2016;15:155-63.
29. Parandeh A, Rahmati-Najarkolaei F, Isfeedvajani MS. Health literacy knowledge and experience survey: Cross-cultural adaptation and the psychometric properties of the Iranian nurse version. J Educ Health Promot 2020;9:244.
30. Mirbagher Ajorpaz N, Zagheri Tafreshi M, Mohtashami J, Zayeri F, Rahemi Z. The effect of mentoring on clinical perioperative competence in operating room nursing students. J Clin Nurs 2016;25:1319-25.
31. Hasty BN, Lau JN, Tekian A, Miller SE, Shipper ES, Bereknyei Merrell S, *et al*. Validity evidence for a knowledge assessment tool for a mastery learning scrub training curriculum. Acad Med 2020;95:129-35.