Yıldız Çelik and Uğur *BMC Medical Education* (2025) 25:919
https://doi.org/10.1186/s12909-025-07504-0

BMC Medical Education

RESEARCH Open Access

Turkish validity and reliability study of the simulation self-report cognitive load measurement tool 2.0

Hilal Yıldız Çelik[1,2] and Esra Uğur[1,2*]

Abstract

Background During simulations, a student's effort to perform the multiple tasks expected of them and the emotional burden arising from the nature of the simulation can create an excessive cognitive load for the student, negatively affecting learning results. Understanding and managing this cognitive load is crucial because it affects how well students can process information and achieve learning objectives.

Objective The aim of this study was to collect evidence of the validity and reliability of the use of the Turkish-adapted Simulation Self-Report Cognitive Load Measurement Tool 2.0, which measures three aspects of cognitive load in nursing students during simulation-based experiences.

Design and methods This was a cross-sectional methodological study of instrument translation and validation. The sample of the study consisted of 101 students in the second year of a nursing undergraduate program at a foundation university. The Turkish-adapted version of the tool was used for data collection. It measures intrinsic cognitive load (five items), extraneous cognitive load (five items), germane cognitive load (five items), and overall cognitive load (one item). Language validity was ascertained via ISPOR's ten-step translation and cultural adaptation guide. Expert opinion was obtained for content validity, and a confirmatory factor analysis for construct validity was performed. Cronbach's alpha coefficient, the split-half method, and item analysis were used to assess the reliability of the tool, and Pearson correlation analysis was used to evaluate relationships between items. Significance was accepted at $p < 0.05$.

Results The tool measures a three-factor structure—intrinsic load, extraneous load, and germane load—and consists of 16 items. In the confirmatory factor analysis, all the factor loadings were greater than 0.30. The fit indices for the model were calculated as $\chi 2/df = 1.423$, with a root mean square error of approximation of 0.0643; the comparative fit index was 0.950; and the standardized root mean square residual was 0.084. The Cronbach's alpha for the tool was 0.822.

Conclusion The analysis provided evidence supporting the validity and reliability of the Turkish-adapted version of the tool for use with a Turkish sample. The use of this tool can aid in developing and enhancing strategies to achieve learning objectives in the design and implementation of clinical simulations.

Clinical trial number Not applicable.

*Correspondence:
Esra Uğur
esra.ugur@acibadem.edu.tr

Full list of author information is available at the end of the article

Keywords Simulation training, Cognitive load, Validity and reliability, Nursing students

Introduction and background

As a teaching strategy in nursing education, simulation provides a learning experience that supports students in developing their psychomotor skills and clinical decision-making abilities in a safe environment [1]. During a simulation, the student is expected to perform multiple tasks, such as applying knowledge, skills, and clinical decision-making [2]. When the emotional load associated with the nature of the simulation is added to the student's effort to meet the learning expectation, the student takes on an excessive cognitive load [2]. An increased cognitive load could negatively affect learning outcomes [3]. Programs that aim to provide skills and competence in a specific subject must therefore be evidence-based, structured in line with learning theories, and designed to provide repeated practice opportunities. One of the learning theories that can help increase the effectiveness of simulation-based learning is the cognitive load theory [4].

Cognitive load theory, developed by Sweller in 1988, provides a conceptual framework for understanding student learning processes and can be used as a guide for teaching procedural skills in a simulated environment [5]. The theory focuses on techniques for managing working memory load during the learning process and for developing teaching methods that enable learners to use their available working memory capacity effectively. According to the theory, individuals use their "working memory," which has limited capacity, and their "long-term memory," which has unlimited capacity, when processing information [6]. Knowledge, skills, or both are stored in long-term memory as hierarchical networks of knowledge described as "schemas" [7]. Before schemas are created in memory, information is processed by working memory, which has the capacity to address only four to six new information items at a time [8].

Two main types of cognitive load—internal and extraneous loads—affect working memory during learning [9]. The "intrinsic load" varies depending on the learner's previous level of expertise and the complexity of the information learned. During the explanation of a challenging topic, an individual who has seen the topic before makes less use of working memory by instead using their existing cognitive schemas. However, when an individual without prior knowledge encounters the same subject, a greater burden is placed on that individual's working memory capacity because of their lack of applicable cognitive schemas [9, 10]. The "extraneous load" is the load on working memory caused by the instructional design used to transfer information. Poorly designed learning materials and teaching methods increase the load on working memory and reduce the resources available for memory learning [8, 10]. Excessive extraneous loads can cause split-attention and redundancy effects. In the split-attention effect, students divide their attention across multiple sources of information and must then combine those sources to solve the problem. In the redundancy effect, students are presented with the same information more than once. The increased load on working memory caused by the split-attention and redundancy effects reduces learning. For this reason, educational materials and methods should be designed to reduce extraneous loads [10].

In long-term memory, another cognitive load defined as the "germane load" ensures that the information sent from working memory to be stored is suitable for learning [10, 11]. Under a germane load, information is processed so that learning can actually occur. It represents the burden imposed by the learner's conscious use of cognitive strategies to organize information for storage in long-term memory [10]. This load has a positive effect on learning and is necessary for learning to occur. To maximize the germane load, an instructional design should assist the student in creating and automating cognitive schemas [12].

For effective learning to occur, the total load created by the combination of the intrinsic, extraneous, and germane loads should not exceed the working memory capacity [10]. Educators who want students to achieve learning outcomes could find that reviewing the cognitive load might lead to a better understanding of the effects of simulation design on those students [3]. The prebriefing standard, which is the best-practice standard published by the International Nursing Association for Clinical Simulation and Learning, states that the cognitive load in a structured prebriefing can increase the effectiveness of a simulation-based experience [13].

Cognitive load theory is used by researchers to understand and manage the cognitive load associated with simulation-based practices. This theory aims to increase students' learning efficiency and academic performance by optimizing their information processing in complex clinical scenarios [14–17]. Simulation experience in nursing education is considered an important learning tool for students to develop their professional skills. However, the cognitive load that students are exposed to during the simulation experience needs to be accurately assessed. In this context, there is a continuing need for tools that can measure the cognitive load experienced by nursing students during simulation in a valid and reliable manner. To address this gap, Josephsen (2018) developed the Simulation Self-Report Cognitive Load Measurement Tool, a cognitive load measurement tool for nursing simulations

[18]. This measurement tool was adapted to nursing simulation experiences from a valid and reliable scale developed by Leppink et al. (2013) to measure intrinsic, germane, and extrinsic cognitive load [19]. The scale was translated into English and Korean and used [16, 18, 20].

Although Leppink et al. (2013) obtained evidence supporting the three-factor structure consisting of intrinsic, extrinsic, and germane cognitive loads using factor analysis, the measurement tool adapted by Josephsen has not been subjected to comprehensive psychometric testing to date. However, other researchers who have used this tool in their studies have reported that the scale demonstrates acceptable to good reliability, with Cronbach's alpha coefficients ranging from 0.67 to 0.92, providing evidence to support its consistent use across different contexts [16, 20]. There is no Turkish version of this tool yet. This cross-sectional methodological study aimed to adapt the tool into Turkish and to gather validity evidence and reliability data to support its use. The questionnaire was specifically adapted for this study and has not been published elsewhere. The English version of the questionnaire is provided as a supplementary file [18].

The models used in validity and reliability studies have evolved over time, reflecting different approaches and frameworks. These models aim to gather evidence to support the use of measurement tools in specific contexts. They can be broadly categorized into two main groups: classical models and modern approaches. Classical models have served as the basic framework for validity and reliability studies for many years. These models typically categorize validity into different types (e.g., content validity, construct validity, criterion validity) and gather evidence for each type separately [21–23]. For reliability, methods such as internal consistency (e.g., Cronbach's alpha) and test-retest reliability are emphasized. Classical models remain valuable because of their practicality and applicability in specific contexts [24]. In contrast, modern approaches, such as Messick's unified validity framework, take a more holistic perspective. Messick (1989) conceptualizes validity as an ongoing process rather than a static property and emphasizes the collection of evidence from five key sources: content, response process, internal structure, relationships with other variables, and consequences. This framework focuses not only on what a tool measures but also on how the results are interpreted and used. This approach is critical to ensuring fair, reliable and ethical assessments, particularly in high-stakes contexts such as health profession education [21, 23].

In this study, classical models were adopted because of their effectiveness in addressing certain aspects of validity and reliability, such as internal structure and relationships with other variables. As noted by Borgersen et al. (2018), certain elements of the classical model can be partially integrated into contemporary frameworks, making it a valuable approach. However, future research may benefit from adopting Messick's unified validity framework to provide a broader and more comprehensive assessment of validity evidence.

Methods

Study design and aim

The purpose of this cross-sectional methodological study was to adapt the Simulation Self-Report Cognitive Load Measurement Tool 2.0 into Turkish and to gather evidence supporting the reliability and validity of the adapted scale. Answers to these research questions were therefore sought:

- Does the Turkish-adapted Simulation Self-Report Cognitive Load Measurement Tool 2.0 provide sufficient validity evidence to support its use in a sample of nursing students?
- Does the Turkish-adapted Simulation Self-Report Cognitive Load Measurement Tool 2.0 demonstrate sufficient reliability evidence in a sample of nursing students?

Participants and setting

The study population consisted of nursing students who were enrolled at the university where the study was conducted and who had prior simulation experience. First-year nursing students were excluded because they had not yet participated in simulation activities during the study period, ensuring that the sample consisted of participants with relevant experience. Simulation-based practices are intensively applied starting in the second year of the undergraduate nursing program. In addition, eliminating the effects that may arise from students' varying scenarios and experiences and ensuring a homogeneous group in terms of cognitive load among the students included in the sample were taken into account; moreover, it was considered that collecting evidence from different students in different groups might influence the process. The study sample initially included 127 s-year nursing students with an active simulation plan during the data collection period. However, students who did not attend school on the day of the scenario ($n=2$), those who declined to participate in the study ($n=4$), and the 10 students who participated in the pilot study to assess the scale's applicability were excluded from the sample.

To ensure an adequate sample size and support the collection of validity and reliability evidence, it was determined that 5 to 20 times the number of participants as items in the tool would be needed [25]. Since the tool consists of 16 items, data from at least 80 students were needed. Ultimately, data collection was completed with the participation of 101 nursing students who agreed to take part in the study, providing sufficient data to evaluate the tool's validity and reliability in line with contemporary frameworks.

Validity evidence

Data were collected via a sociodemographic form and the validated Turkish-adapted Simulation Self-Report Cognitive Load Measurement Tool 2.0. The socio-demographic form developed by the researchers for descriptive purposes recorded each participant's age, gender and year group, and was specifically designed to include only these questions.

Simulation self-report cognitive load measurement tool 2.0

The Simulation Self-Report Cognitive Load Measurement Tool 2.0 was developed by Pass et al. (2013) to measure three different cognitive loads accurately. Although this tool helps educators better understand the differences between students or learning outcomes achieved through instructional design, it also measures potential environmental and instructional design hindrances to student learning [19]. The tool was subsequently adapted by Josephen (2018) to measure the cognitive loads experienced by students during simulation experiences. The tool's 16 items measure intrinsic load (five items), extraneous load (five items), germane load (five items), and overall cognitive load (one item) [18]. The first 12 items are scored between 0 and 10 (0, not at all true; 10, completely true), whereas items 13–16 are scored between 0 and 9 (items 13: 1, very, very low mental effort; 9, very, very high mental effort; items 14 and 15: 1, very, very easy; 9, very, very difficult; item 16: 1, a lot, very little; 9, a lot, a lot)]. The original tool contains no negative statements. The mean total score on the tool indicates the cognitive load; an increasing score indicates an increasing cognitive load. The Cronbach's alpha values calculated for the items measuring intrinsic, extraneous, and germane cognitive loads, as well as the overall score, were found to be 0.775, 0.358, 0.841, and 0.736, respectively [18].

Translation

The 10-step translation and cultural adaptation guide published by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) was used to ensure the language validity of the tool [26]. The work carried out in the ten steps was as follows:

Preparation First, permission was obtained from the tool owner to avoid name or copyright infringement. The studies carried out by the tool owner on the subject were investigated.

Translating the scale into Turkish The scale was independently translated into Turkish by a faculty member living in the United States who is a linguist fluent in both languages.

Reconciliation The original was compared with the translation to check for ambiguous expressions and inconsistencies that might have occurred between languages. The translation was considered complete when all the items agreed. A check for spelling errors was also performed.

Back translation The Turkish form of the tool was independently translated back into English by a native English speaker who is a linguist living in Turkey, and the consistency of the two English forms of the tool was evaluated.

Review of back translation The translated tool was sent to the tool owner for an evaluation of its meaning and similarity. The Turkish version was found to be faithfully compatible with the original tool.

Adaptation The translated tool was presented for an assessment of content validity to ten faculty members who had completed their doctorates in various fields of nursing and who provided simulation-based education. These experts were asked to evaluate each item on a 1- to 4-point scale (1, not appropriate; 2, the item requires a change to an appropriate form; 3, appropriate, but minor changes needed; 4, very appropriate) in terms of suitability for purpose and understandability. Their opinions on improving the items were also sought.

Cognitive analysis The adapted tool was administered to 10 students who agreed to participate in a pilot study. The students who participated in the pilot study received a brief, clear summary of the purpose of the research and tool administration. The students then provided oral and written feedback. During the pilot application of the scale, students stated that they had difficulty using the 10-point scale when evaluating the items. Considering that cultural differences may be effective, the 10-point Likert-type rating was reduced to a 5-point Likert-type. Items 13–16 of the measurement tool are scored between 0 and 5 (item 13: 1, quite low; 5, quite high mental effort; items 14 and 15: 1, quite easy; 5, quite difficult; item 16: 1, quite little; 5, quite a lot).

Evaluation after cognitive analysis After the pilot study, the wording of the items was edited in line with the feedback received from the participating students.

Final reading The adapted tool was given to an instructor who had not seen any version of it before it was checked for spelling, punctuation, and reading errors.

Final reporting All the data were recorded during the phases of the language and cultural adaptation of the tool, and the tool was then made ready to measure the student cognitive loads during the simulations.

Data collection process

Evidence for validity and reliability was gathered between April and December 2022. As part of the Internal Medicine Nursing course, after a theoretical background provided about the topic, the students who made up the sample group in this study took part in a scenario application that required them to make clinical decisions for a patient experiencing respiratory distress. The simulation experience was designed and conducted by the faculty members responsible for the pertinent course in accordance with the Healthcare Simulation Standards of Best Practice™ developed by the International Nursing Association of Clinical Simulation and Learning (INACSL) [27].

The simulation experience consisted of a prebriefing, an experience phase, and debriefing sessions. Each scenario lasted roughly ten minutes, and students took part in it in groups of two. In the simulation scenario while one student took on the role of a nurse, the other took on the role of a student nurse. After every ten to eight students' scenario experience, they were taken to a 30- to 40-minute group debriefing session. In this session, the "Promoting Excellence and Reflective Learning in Simulation" method was applied [28]. At the end of the debriefing session, the students were asked to fill out the "Simulation Self-Report Cognitive Load Measurement Tool 2.0" scale in the debriefing room by researchers. To maintain the integrity of the response process immediately after the debriefing sessions, the questionnaire was completed by each student in a controlled and distraction-free environment.

To ensure the quality and reliability of the response process, participants were provided with clear and detailed information about the study's objectives, their role, and the data collection procedures. Before completing the tool, the participants received a brief explanation to clarify how to use the tool, and the researchers were present to address any questions. The tool was administered in a quiet, distraction-free setting immediately after the simulation-based practices to minimize recall bias and accurately capture participants' experiences. The participants completed the tool individually to maintain the authenticity of their responses. These measures collectively ensured that the data collected reflected the intended constructs and maintained the integrity of the study's findings.

Data analysis

The study data were analyzed using IBM SPSS Statistics (version 24.0) and Amos software (IBM, Acıbadem University, Turkey). Categorical variables were summarized using frequencies and percentages, while numerical variables were described using descriptive statistics, including mean, standard deviation, median, minimum, and maximum values. The validation process was designed to provide evidence for the validity of the adapted tool across multiple dimensions. Content validity was evaluated to ensure that the items adequately represented the intended constructs. Item analyses, including item–total correlations, were conducted to assess the contribution of individual items to the overall structure of the tool. Confirmatory factor analysis (CFA) was performed to evaluate the internal structure of the tool and its alignment with the theoretical framework of the original version. Reliability was assessed through internal consistency (Cronbach's alpha coefficients), the split-half method, and item analyses, providing evidence for the consistency of the tool in measuring the intended constructs. Pearson correlation analysis was conducted to evaluate the relationships between items and subdimensions, contributing to the understanding of the tool's internal structure and its ability to measure the targeted constructs. Statistical significance was set at $p < 0.05$. The authors adhered to the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines throughout the study, ensuring a rigorous and systematic approach to the validation process [29].

Ethics considerations

Permission via e-mail was received from the author of the tool to adapt it to Turkish and to conduct validity and reliability studies. Before data collection began, ethics approval for the study was obtained from the Medical Research Ethics Committee of the institution where the study was conducted (no. 2022-03/17 dated 14 February 2022), and written institutional permission was obtained from the university administration. The research was conducted in accordance with the Declaration of Helsinki, the institutional Code of Ethics, and complied with all relevant guidelines and regulations. The study's data collection phase was carried out during the academic year, following the Internal Medicine Nursing course's simulation experience. The research group did not involve the course instructors, and the study was carried out entirely apart from the course. Prior to the data collection phase, the students were given comprehensive information by the researchers regarding the goals, procedures, and scope of the study. They were also informed that participation was entirely voluntary, that they could leave the study at any time without facing any repercussions, and that confidentiality and anonymity would be carefully maintained. The students were also made fully aware that their course success evaluations would not be impacted by the data collection procedure. Students who agreed to participate in the study signed informed consent forms before participating in the process.

Results

The average age of the student participants was 21.98 ± 1.17 years (min–max: 19–28 years).

Psychometric measurements

Content evidence

After the translation phase was complete, the translated tool was presented to 10 academicians, experts in the field, to test content validity. The experts assessed the relevance, clarity, and representativeness of each item in relation to the intended constructs. The Davis technique was used to calculate the content validity index [29], which was determined to be 0.80. The content validity ratio was determined to be 1.00.Item Analysis.

The Cronbach's alpha coefficient was used to test the reliability of the individual subdimensions, providing evidence for the internal consistency of the tool. To determine the contribution of each item to the overall reliability, the "alphas if the item was deleted" values were calculated. These values indicate the internal consistency of the remaining items if any individual item is removed, offering insights into the structural integrity of the tool. For the adapted tool, a reliability value of $\alpha = 0.822$ was obtained, demonstrating acceptable internal consistency. Table 1 presents the effects of the individual items on reliability.

The item–total correlations ranged between 0.328 and 0.719, with the highest correlations observed for items 11 and 10, and the lowest correlation observed for item 7. These correlations provide evidence that the majority of the items are strongly aligned with the overall construct being measured. Table 2 presents the item analysis for the tool, revealing that the scores for all the questions in the adapted tool were statistically significantly higher for the upper 27% of the participants than for the lower 27% ($p = 0.001$ and $p < 0.01$, respectively). This finding supports the tool's ability to differentiate between participants with higher and lower levels of the construct being measured, further contributing to its validity evidence.

Internal structure evidence

Confirmatory factor analysis (CFA) was conducted to evaluate the internal structure of the Turkish-adapted tool and to determine the alignment of its items and subdimensions with the original structure. Figure 1 presents the standardized factor loadings of the items forming the three subdimensions of the adapted tool, as identified through discriminant function analysis. Item 13 was not included in the subdimension structure but was retained for inclusion in the total score. The highest factor loading was observed for item 11, followed by items 10 and 14, indicating that these items strongly contribute to the measurement of the intended construct.

To assess the model fit, several fit indices were calculated. The fit criterion of the model was $\chi 2/df = 1.423$, the root mean square error of approximation (RMSEA) was 0.0643, the comparative fit index (CFI) was 0.950, and the standardized root mean square residual (SRMR) was 0.084 (Table 3). These values indicate an acceptable to good fit of the model to the data, providing evidence that

Table 1 Effects on the reliability of items constituting factors of the Turkish-adapted simulation Self-Report cognitive load measurement tool 2.0

	When the item is deleted		Corrected item–total correlation	Cronbach alpha when the item is deleted	Item–total test correlation
	Scale average	Scale variance			
Item 1	52.7798	297.025	0.557	0.807	0.599
Item 2	52.6514	300.155	0.548	0.807	0.658
Item 3	52.7706	309.345	0.449	0.816	0.551
Item 4	53.8624	306,120	0.577	0.805	0.522
Item 5	52.3670	285.457	0.625	0.801	0.453
Item 6	54.9633	354.313	0,110	0.833	0.399
Item 7	51.8532	314.626	0.418	0.818	0.328
Item 8	55.2018	361.237	0.342	0.835	0.393
Item 9	54.9908	328.472	0.536	0.811	0.546
Item 10	55.0734	319.309	0.658	0.804	0.715
Item 11	55.0459	318.729	0.681	0.803	0.719
Item 12	54.5505	314.453	0.568	0.807	0.598
Item 13	53.8532	355.349	0.442	0.822	0.399
Item 14	54.1743	348.330	0.602	0.817	0.591
Item 15	54.7339	342.271	0.671	0.814	0.585
Item 16	53.8349	375.621	0.255	0.836	0.383

Table 2 Item analysis of the Turkish-adapted simulation Self-Report cognitive load measurement tool 2.0

	Lower 27%		Upper 27%		t	p Value[a]
	Mean ± SD	Median (min–max)	Mean ± SD	Median (min–max)		
Item 1	2.07 ± 2.45	1 (0–8)	8.05 ± 2.55	9 (2–10)	8.208	0.001
Item 2	2.23 ± 2.62	1.5 (0–9)	7.95 ± 2.22	9 (3–10)	7.869	0.001
Item 3	2.00 ± 2.30	2 (0–9)	7.05 ± 2.66	7 (2–10)	7.048	0.001
Item 4	8.87 ± 1.48	10 (5–10)	3.89 ± 2.79	4 (0–10)	8.154	0.001
Item 5	1.87 ± 2.00	1 (0–7)	8.63 ± 1.34	9 (6–10)	13.007	0.001
Item 6	6.40 ± 2.37	7 (2–10)	8.58 ± 2.95	9 (2–10)	3.073	0.005
Item 7	6.70 ± 2.18	6.5 (2–10)	2.42 ± 2.43	3 (0–10)	6.393	0.001
Item 8	6.37 ± 1.87	6.5 (3–10)	7.74 ± 2.9	8 (0–10)	2.220	0.025
Item 9	9.33 ± 1.56	10 (4–10)	5.47 ± 2.2	5 (0–9)	7.193	0.001
Item 10	9.53 ± 0.94	10 (6–10)	5.11 ± 2.47	5 (0–10)	8.903	0.001
Item 11	9.53 ± 0.86	10 (7–10)	4.84 ± 1.86	5 (0–8)	11.696	0.001
Item 12	9.40 ± 1.10	10 (6–10)	5.11 ± 3.28	5 (1–10)	6.636	0.001
Item 13	3.33 ± 0.55	3 (2–4)	4.26 ± 0.81	4 (3–5)	4.426	0.001
Item 14	2.67 ± 0.66	3 (1–4)	4.16 ± 0.76	4 (3–5)	7.240	0.001
Item 15	2.00 ± 0.79	2 (1–3)	3.89 ± 0.74	4 (3–5)	8.404	0.001
Item 16	3.97 ± 0.72	4 (2–5)	3.21 ± 0.85	3 (2–5)	3.334	0.002

[a] By *Student's t* test

SD, standard deviation

the internal structure of the adapted tool aligns well with the theoretical framework of the original tool.

Cronbach's alpha analysis was conducted to evaluate the internal consistency of the results obtained from the Turkish-adapted tool, providing evidence for the reliability of the tool. The total Cronbach's alpha for the adapted tool was found to be 0.822, indicating good internal consistency. The Cronbach's alpha values for the intrinsic, extraneous, and germane cognitive load subdimensions were 0.757, 0.572, and 0.848, respectively (Table 4).

The split-half method was used to evaluate the internal consistency of the Turkish-adapted tool, providing evidence for its internal structure. The individual items were divided into two halves: odd and even. The Cronbach's alpha value for the first half was 0.635, and for the second half, it was 0.667. The relationship between the averages of the two halves was 0.687. The associated Spearman–Brown coefficient was 0.800, and the Guttman split-half coefficient was 0.796.

Scale scores of nursing students

The average scores for this student cohort were 21.11 ± 9.84 for the intrinsic subdimension, 17.55 ± 7.21 for the extraneous subdimension, and 34.56 ± 8.24 for the germane subdimension. The average total score was 76.88 ± 11.96.

The relationships between the subdimensions and the total score provide evidence for the validity of the adapted tool. A statistically significant positive correlation ($r = 0.325$) was observed between the scores for the intrinsic and extraneous subdimensions, indicating a moderate relationship between these constructs. A statistically significant negative correlation ($r = -0.192$, $p = 0.001$) was found between the intrinsic and germane subdimensions, suggesting that these constructs may represent distinct aspects of cognitive load.

The positive relationship between the scores for the intrinsic subdimension and the total score of the Turkish-adapted tool was statistically significant ($r = 0.659$; $p < 0.01$), demonstrating that the intrinsic subdimension contributes strongly to the overall construct measured by the tool. Similarly, the positive relationships between the scores for the extraneous subdimension and the total score ($r = 0.562$; $p < 0.01$) and between the scores for the germane subdimension and the total score ($r = 0.520$; $p < 0.01$) further support the internal structure and the alignment of the subdimensions with the overall construct. No statistically significant relationship was detected between the scores for the extraneous and germane subdimensions ($p > 0.05$), which may indicate that these subdimensions measure distinct aspects of cognitive load. These findings collectively provide evidence for the relationships with to other variables, supporting the validity of the inferences drawn from the adapted tool's scores in this specific context (Table 5).

Discussion

In simulation design, measuring and balancing the participants' cognitive load can increase the effectiveness of the simulation-based experience, supporting participants in achieving their learning goals and enabling them to use their attention and information-processing capacity efficiently [13, 18]. An examination of the national and international literature to determine a valid and reliable

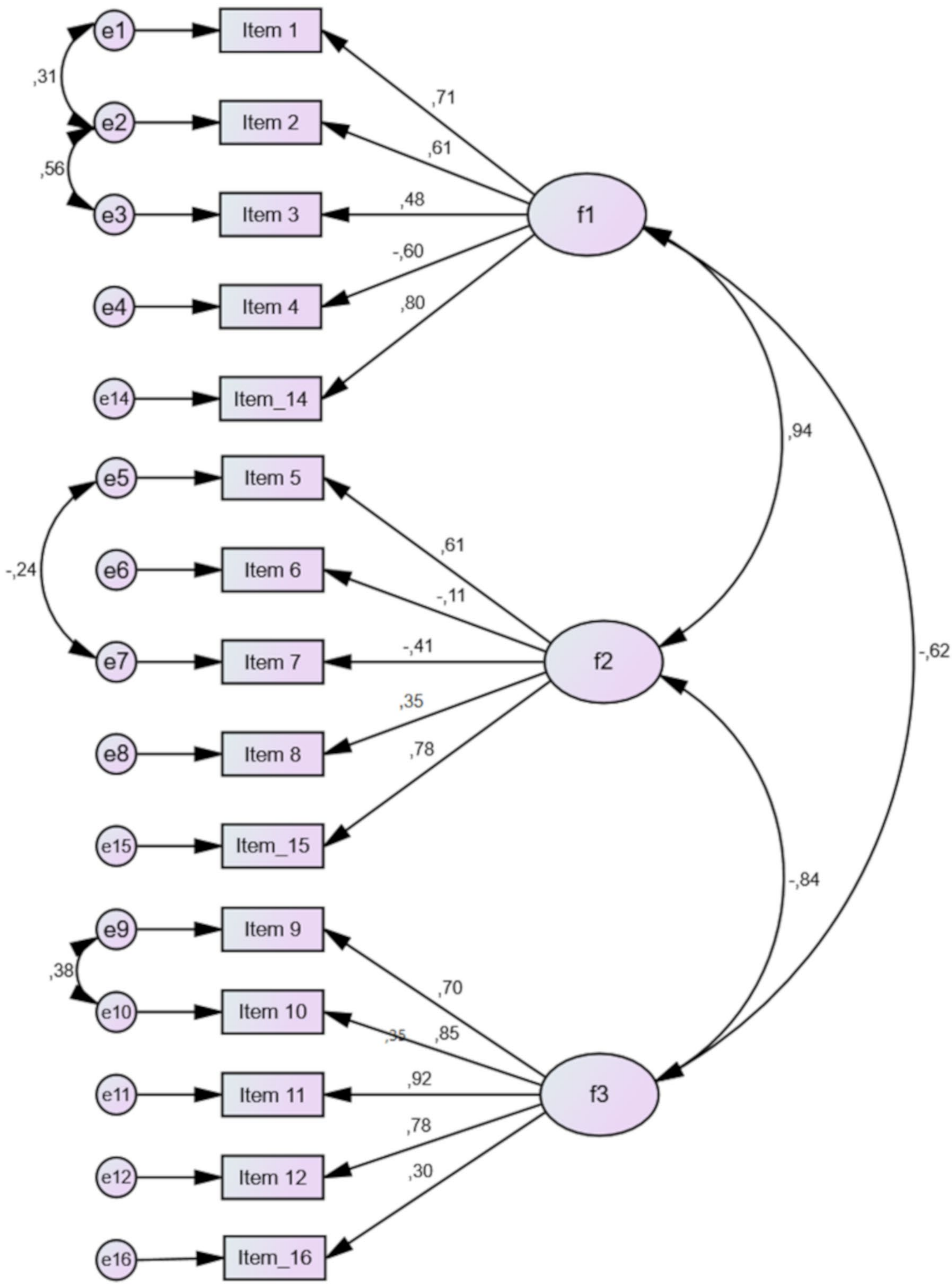

Fig. 1 Confirmatory factor analysis for the Turkish-adapted Simulation Self-Report Cognitive Load Measurement Instrument 2.0

Table 3 Confirmatory factor analysis results of the Turkish-adapted simulation self-reported cognitive load measurement tool 2.0

Criterion	Fit ranges[a]		Result of model	Fit type observed
	Good	Acceptable		
SRMR	$0 \leq SRMR \leq 0.05$	$0.05 \leq SRMR \leq 0.10$	0.084	Good
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 3$	1.423	Good
GFI	$0.95 \leq GFI \leq 1$	$0.90 \leq GFI \leq 0.95$	0.867	—
AGFI	$0.90 \leq AGFI \leq 1$	$0.85 \leq AGFI \leq 0.90$	0.808	—
RFI	$0.90 \leq RFI \leq 1$	$0.85 \leq RFI \leq 0.90$	0.813	—
IFI	$0.97 \leq IFI \leq 1$	$0.95 \leq IFI \leq 0.97$	0.951	Acceptable
TLI	$0.97 \leq NNF \leq 1$	$0.95 \leq NNFI \leq 0.97$	0.936	—
CFI	$0.97 \leq CFI \leq 1$	$0.95 \leq CFI \leq 0.97$	0.950	Acceptable
RMSEA	$0 < RMSEA < 0.05$	$0.05 \leq RMSEA \leq 0.10$	0.0643	Acceptable

[a] (Erkorkmaz et al., 2013)

SRMR, standardized root mean square residual; df, degrees of freedom; GFI, goodness-of-fit index; AGFI, adjusted GFI; RFI, relative fit index; IFI, incremental fit index; TLI, Tucker Lewis index; NNF, nonnormed fit; NNFI, NNF index; CFI, comparative fit index; RMSEA, root mean square error of approximation

Table 4 Internal consistency of the Turkish-adapted simulation self-reported cognitive load measurement tool 2.0

Subdimension	Included items	Min–max (median)	Mean ± SD	Internal consistency		
				Spearman–Brown	Guttman	Cronbach alpha
Intrinsic	1, 2, 3, 4, 14	7–44 (25)	24.25 ± 8.15	0.548	0.626	0.757
Extraneous	5, 6,7, 8,15	17–42 (28)	28.30 ± 4.74	0.654	0.556	0.572
Germane	9, 10, 11, 12, 16	13–45 (36)	34.56 ± 8.24	0.845	0.737	0.848
Total	1–16	43–117 (88)	87.11 ± 12.32	0.779	0.721	0.822

** Spearman–Brown and Guttman are values obtained via split-half methods

SD, standard deviation

Table 5 Correlations between factors

		Factor 1 (intrinsic)	Factor 2 (extraneous)	Factor 3 (germane)
Factor 1 (intrinsic)				
	r[a]	1		
	p	—		
Factor 2 (extraneous)				
	r[a]	0.325		
	p	0.001		
Factor 3 (germane)				
	r[a]	−0.192	0.058	
	p	0.001	0.548	
Total				
	r[a]	0.659	0.562	0.520
	p	0.001	0.001	0.001

[a] By Pearson correlation

tool that measures cognitive load during a student simulation experience led to the choice of the Simulation Self-Report Cognitive Load Measurement Tool 2.0. In the present study, we aimed to adapt the tool for use in Turkish culture and to evaluate the validity of the inferences drawn from its scores, as well as its reliability. There are no studies on the adaptation of the scale to different languages and cultures. Therefore, the results of our adaptation study are discussed only by comparing them with the findings of the original scale.

The content and construct validity of the tool were analyzed to provide evidence for its ability to measure the intended constructs. Content validity was evaluated by consulting experts, who assessed the extent to which each item of the tool adequately represented the concept it was intended to measure [30, 31]. In this study, 10 academics were asked for their opinions. The content validity index obtained from such a consultation varies depending on the number of experts, and when the number of experts consulted is 10, the content validity index achieved must be at least 0.78 [32]. The content validity index (CVI) of all the items in the tool was found to be 0.80, and the content validity ratio (CVR) was 1.00. These results provide evidence that each item, subdimension, and total score of the tool adequately represent the intended constructs, supporting the content validity of the adapted tool.

Confirmatory factor analysis was used to examine the construct validity of the tool. Confirmatory factor analysis verifies the factors determined in measurement tools adapted from another language [33–35]. The indices used in CFA are χ2/df, the goodness-of-fit index (GFI), the incremental fit index (IFI), CFI, RMSEA, and SRMR [36, 37]. The acceptable fit value for the GFI, IFI, and CFI is ≥0.90, and a value of ≥0.95 is considered good. The acceptable fit value for χ2/df is between 3 and 4, with a value between 0 and 3 being considered good. The acceptable fit value for the RMSEA is 0.05–0.08, with a good fit value of 0–0.05. The acceptable fit value for the SRMR is 0.05–0.10, with a good fit value of 0–0.08 [37, 38]. The fit values for the Turkish-adapted tool were χ2/df, 1.423; GFI, 0.867; IFI, 0.951; CFI, 0.950; RMSEA, 0.0643; and SRMR, 0.084. These results indicate good and acceptable fit indices, with a good fit for the corrected chi-square value, supporting the construct validity of the adapted tool ($p = 0.001$; $p < 0{,}01$). The fit values in the study by Leppink et al. were CFI, 0.995; TLI, 0.992; and RMSEA, 0.035 [19]. The differences between the fit criteria for the original tool and the Turkish-adapted tool were minimal and were attributed to linguistic and cultural adaptations made during the translation process. These differences reflect the influence of cultural and contextual factors on the tool's structure while maintaining its overall validity.

The CFA of the tool revealed that the factor loadings were 0.48–0.80 for the intrinsic load subdimension, 0.31–0.78 for the extraneous load subdimension, and 0.30–0.92 for the germane load subdimension. The literature indicates that items with a factor loading of less than 0.30 should be removed from the pertinent dimension [39]. The overall load (item 13) is not included in the dimension structure because this item has a different structure than the others. However, the score for the overall load is added to the total score. The critical values and significance values demonstrated that all the items in the subdimensions were necessary. The present study determined that the factor loadings were not less than 0.30 and were within acceptable limits. In the original study, Leppink et al. (2013) included error correlations between some items in the germane subdimension and stated that this could explain the effects of different teaching styles. Accordingly, the tool can be said to have an acceptable factor structure for the Turkish sample [19].

Reliability is defined as the ability of a tool to measure a desired characteristic consistently [40]. Internal consistency means that all the items and subdimensions measure the same concept. The methods most commonly used to determine the internal consistency of a tool's measurements include the Cronbach's alpha reliability coefficient and the item–total score reliability [25, 41]. The Cronbach's alpha coefficient, which reveals the internal consistency of measurements, is generally quite reliable in the range of 0.60–0.80 and highly reliable in the range of 0.80–1.00 [42]. Accordingly, the extraneous load subdimension of the scale had a low level of reliability (Cronbach's $\alpha = 0.572$), while the intrinsic load (Cronbach's $\alpha = 0.757$) and germane load (Cronbach's $\alpha = 0.848$) subdimensions were found to be quite reliable. When all the items of the scale were evaluated together, Cronbach's α value was calculated as 0.822 and it was determined that the scale had a high level of reliability in general.

In the correlation between the scores for each item and the total tool score, a high correlation coefficient indicates that the relevant item measures the same concept as other items in the test. A low correlation coefficient indicates that the relevant item measures a concept different from those measured by the other items in the test [43, 44]. The item–total score correlation values for the Turkish-adapted tool items ranged between 0.328 and 0.688. All item–total correlation coefficients were 0.30 and greater, as recommended [45]. Notably, this coefficient is significant at the 95% confidence interval and has a significant *p* value [25, 35]. The scores for all the items in the tool were statistically significantly higher for the upper 27% of the participants than for the lower 27% ($p = 0.001$; $p < 0.01$). While the item-total score correlation values for the scale adapted to Turkish vary between 0.328 and 0.688, items 6 and 8 have lower correlation values compared to other items. Although there is no definitive consensus on the acceptable range of the correlation coefficient, it is generally stated in the literature that it should be between 0.30 and 0.50 [43, 46]. In addition, it is emphasized that the removal of items with an item-total score correlation coefficient below 0.30 from the scale should be decided by evaluating the effect of these items on the Cronbach's alpha coefficient [35, 47]. This study determined that items 6 and 8 would not be deleted from the scale, as their exclusion did not result in a substantial change in the Cronbach's alpha value. The item-total score correlation coefficients of the other items are at satisfactory levels regarding reliability.

The observation of a statistically significant negative correlation (r = -0.192, $p < 0.001$) between the "intrinsic" and "germane" subdimensions is an unexpected finding and points to problems in defining the concepts or to specific effects in the context of the Turkish sample. Theoretically, a positive relationship between the intrinsic load and the germane load is expected, but this negative correlation suggests that intrinsic load may hinder cognitive processes rather than support the germane load. This result can be evaluated in the context of individual differences. Although the general intrinsic load average of the group is at a moderate level, students with lower prior knowledge may have difficulty in creating germane load by perceiving intrinsic load higher. This situation shows that individual prior knowledge levels may affect the relationship between intrinsic and germane load. It is also known that the scales used to measure different subdimensions of cognitive load do not always work perfectly [19]. These results indicate that the Turkish-adapted tool is strong in terms of both content validity and construct validity and is suitable for making a reliable assessment of the cognitive load of nursing students.

The split-half method was used to test the reliability of the measurements obtained via the Turkish-adapted tool. A high reliability coefficient indicates that the original and Turkish-adapted tools are both reliable; an insufficiently high coefficient indicates that the reliability of both tools is low [48]. Because no other scale that measures the three types of cognitive load in simulation-based education within nursing education is available, the criterion-related validity method was not employed. The Cronbach's alpha for the first half was 0.635, and for the second half, it was 0.667. The relationship between the averages of the two halves was 0.687, indicating a moderate level of consistency between the halves and supporting the internal structure of the tool. Additionally, the Spearman–Brown coefficient of 0.800 and the Guttman split-half coefficient of 0.796 provide evidence for the internal consistency and reliability of the tool, supporting its ability to measure the intended constructs with a high degree of precision.

Our results demonstrate that the Turkish-adapted tool is valid and reliable for measuring three different cognitive loads in nursing students during simulation practices. Researchers can therefore use this tool in relevant research. The tool is recommended for use, especially in students with simulation experience; however, its use in students with no previous simulation experience might be premature given the anxiety and stress specific to a first simulation practice. These factors must therefore be taken into consideration during the application of the tool.

Limitations

The inclusion of second-year nursing students from a single foundation university in Turkey should be considered when generalizing the findings of this study. This tool is a self-assessment instrument, and students may vary in their ability to perform accurate self-evaluations. Such variability could influence the reliability and validity of the results. The findings may be influenced by differences in students' prior knowledge levels, the instructional design of the simulation, and the self-reported nature of the tool. Variations in prior knowledge among students could lead to inconsistent perceptions of extraneous cognitive load, while differences in instructional design may have shaped how students experienced cognitive demands during the simulations. Additionally, as a self-report tool, individual differences in how students perceive and interpret their cognitive load may contribute to response variability, potentially affecting internal consistency. Although the sample size ($n = 101$) was sufficient for statistical analyses, the fact that participants were drawn from a single institution may limit the generalizability of the findings. Differences in curriculum, instructional methods, and student demographics across institutions could further influence cognitive load perceptions. These contextual factors underscore the importance of interpreting reliability and validity estimates within the specific context of the study.

Future research with larger and more diverse student samples is recommended to enhance external validity and ensure broader applicability of the results. Such studies could provide additional evidence for the tool's validity and reliability across different educational settings and student populations.

Conclusion and recommendations

The Turkish-adapted Simulation Self-Report Cognitive Load Measurement Tool 2.0 demonstrated acceptable psychometric properties, providing evidence supporting its use in assessing the cognitive load of nursing students during simulation practices. The tool consists of three factors and 16 items, encompassing intrinsic, extraneous, and germane cognitive load, as well as total cognitive load. On the basis of these findings, the Turkish-adapted tool is recommended for measuring various aspects of cognitive load among nursing students engaged in simulation practices.

The validation process was guided by a framework that emphasizes the importance of evaluating the inferences drawn from the tool's scores rather than the tool itself. This approach ensures that the tool's validity is context-specific and supported by evidence from multiple sources, including internal structure and relationships between subdimensions. This study, within the context of external evidence, provides additional evidence for the validity of the tool developed to measure cognitive load and support the development of cognitive load theory and contributes to making the model more comprehensive and universally applicable. Future studies should consider correlating the tool's scores with external indices, such as objective measures of cognitive load, to strengthen the validity evidence. Although the sample size ($n = 101$) was sufficient for the statistical analyses performed, the fact that participants were selected from a single foundation university in Turkiye may limit the generalizability of the findings. Differences in curriculum, teaching methods, and student demographics across institutions may influence students' perceptions of cognitive load. To enhance external validity and ensure broader applicability of the results, future studies should test the tool with larger and more diverse sample groups.

The use of this tool may contribute to the development of strategies aimed at improving the design and implementation of simulation practices in nursing education. By assessing students' perceptions of cognitive load during simulation applications, the tool can help identify areas where students may experience excessive cognitive strain, enabling instructors to adapt their teaching strategies accordingly. Additionally, the tool can be used to evaluate the effectiveness of different simulation-based teaching methods, optimize scenario complexity, and enhance the overall learning experience of nursing students.

Abbreviations

ISPOR	International society for pharmacoeconomics and outcomes research
INACSL	International nursing association of clinical simulation and learning
COSMIN	Consensus-based standards for the selection of health measurement instruments
RMSEA	Root mean square error of approximation
CFI	Comparative fit index
SRMR	Standardized root mean square residual
IFI	Incremental fit index
GFI	Goodness of fit index
CFA	Confirmatory factor analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07504-0.

Supplementary Material 1

Acknowledgements

This study was supported by Scientific and Technological Research Council of Turkey (TUBITAK) under the Grant Number 222S743. The authors thank to TUBITAK for their supports.

Author contributions

HYÇ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing–original draft, Writing–review & editing. EU: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing–original draft, Writing–review & editing. All authors approved the final manuscript for submission.

Funding

This study was supported by Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 222S743. The authors thank TUBITAK for their supports.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Medical Research Ethics Committee of Acıbadem University and Acıbadem Healthcare Institutions (Date: 14.02.2022 & No: ATADEK-2022-03/17). Permission was obtained from the universities. Informed consent was obtained from those who agreed to participate in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Other acknowledgments

The authors sincerely thank the participants their contribution. Also, the authors would like to thank and the administrations at the universities where the data were collected for their support.

Author details

[1]Department of Nursing, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye
[2]Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye

Received: 28 August 2024 / Accepted: 6 June 2025
Published online: 01 July 2025

References

1. Harris MA, Pittiglio L, Newton SE, Moore G. Using simulation to improve the medication administration skills of undergraduate nursing students. Nurs Educ Perspect. 2014;35(1):26–9. https://doi.org/10.5480/11-552.1
2. Fraser K, Ma I, Teteris E, Baxter H, Wright B, Mclaughlin K. Emotion, cognitive load and learning outcomes during simulation training. Med Educ. 2012;46(11):1055–62. https://doi.org/10.1111/J.1365-2923.2012.04355.X
3. Rogers BA, Franklin AE. Cognitive load experienced by nurses in simulation-based learning experiences: an integrative review. Nurse Educ Today. 2021;99:104815.
4. Mcdonald C, Davis M. Benson · cole. Using Evidence-Based learning theories to guide the development of virtual simulations. Clin Soc Work J. 2021;49:197–206. https://doi.org/10.1007/s10615-021-00809-9
5. Haji FA, Cheung JJH, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. Thrive or overload? The effect of task complexity on novices' simulation-based learning. Med Educ. 2016;50(9):955–68. https://doi.org/10.1111/MEDU.13086
6. Sweller J. Discussion of the special issue on cognitive load theory. Br J Educ Psychol. 2023;93:402–10. https://doi.org/10.1111/bjep.12606.
7. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010;44(1):85–93. https://doi.org/10.1111/J.1365-2923.2009.03498.X
8. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. Perspect Med Educ. 2015;4:107–9. https://doi.org/10.1007/s40037-015-0193-9
9. Sweller J. Cognitive load theory. Psychol Learn Motivation - Adv Res Theory. 2011;55:37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8
10. Sweller J. Cognitive load theory and educational technology. Education Tech Research Dev. 2020;68(1):1–16. https://doi.org/10.1007/S11423-019-09701-3/TABLES/1
11. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. Educ Psychol. 2003;38(1):1–4. https://doi.org/10.1207/S15326985EP3801_1. /ASSET//CMS/ASSET/CCCD6D20-9336-4E97-9C0F-0DC7631208E0/S15326985EP3801_1.FP.PNG.
12. Josephsen J. Cognitive load theory and nursing simulation: an integrative review. Clin Simul Nurs. 2015;11(5):259–67. https://doi.org/10.1016/j.ecns.2015.02.004
13. McDermott DS, Ludlow J, Horsley E, Meakim C. Healthcare simulation standards of best practicetm prebriefing: Preparation and briefing. Clin Simul Nurs. 2021;58:9–13.

14. Danielson JA, Eric MM, Pamela JV, Vanessa AP, Karen MY, Mary MC, Jeanne WG, Holly RDW. Characteristics of a cognitive tool that helps students learn diagnostic problem solving. Education Tech Research Dev. 2007;55:499–520. https://doi.org/10.1007/s11423-006-9003-8
15. Faber TJ, Dankbaar ME, van den Broek WW, Bruinink LJ, Hogeveen M, van Merriënboer JJ. Effects of adaptive scaffolding on performance, cognitive load and engagement in game-based learning: a randomized controlled trial. BMC Med Educ. 2024;24(1):943. https://doi.org/10.1186/s12909-024-05698-3
16. Jung MJ, Roh YS. Mediating effects of cognitive load on the relationship between learning flow and clinical reasoning skills in virtual simulation learning. Clin Simul Nurs. 2022;64:16–23. https://doi.org/10.1016/j.ecns.2021.12.004
17. Mayrath MC, Nihalani PK, Robinson DH. Varying tutorial modality and interface restriction to maximize transfer in a complex simulation environment. J Educ Psychol. 2011;103(2):257. https://doi.org/10.1037/a0022369
18. Josephsen J. Cognitive load measurement, Worked-Out modeling, and simulation. Clin Simul Nurs. 2018;23:10–5. https://doi.org/10.1016/j.ecns.2018.07.004
19. Leppink J, Paas F, Van der Vleuten M, Van Gog CP, Van Merriënboer TG. Development of an instrument for measuring different types of cognitive load. Published Online. 2013. https://doi.org/10.3758/s13428-013-0334-1
20. Hee NY, Young SR. Effects of Peer-Led debriefing on cognitive load, achievement emotions, and nursing performance. Clin Simul Nurs. 2022;55:1–9. https://doi.org/10.1016/j.ecns.2021.03.008
21. Messick S. Meaning and values in test validation: the science and ethics of assessment. Educational Researcher. 1989;18(2):5–11. https://doi.org/10.3102/0013189X018002005
22. Hamstra SJ, Yamazaki K. A validity framework for effective analysis and interpretation of milestones data. J Graduate Med Educ. 2021;13(2 Suppl):75–80. https://doi.org/10.4300/JGME-D-20-01039.1
23. Roach VA, Validity. Conceptualizations for anatomy and health professions educators. Anat Sci Educ. 2025;00:1–6. https://doi.org/10.1002/ase.70016
24. Borgersen NJ, Naur TM, Sørensen SM, Bjerrum F, Konge L, Subhi Y, Thomsen AS. S. Gathering validity evidence for surgical simulation: a systematic review. Ann Surg. 2018;267(6):1063–8. https://doi.org/10.1097/SLA.0000000000002652
25. Tavşancıl E. Measuring attitudes and data analysis with SPSS. Tutumların ölçülmesi ve spss ile veri analizi Ankara: Nobel Yayınları. Published online 2014.
26. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for Patient-Reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation Backg. Value Health. 2005;8(2):94–104. http://www.ispor.org. Accessed November 3, 2023.
27. Watts PI, McDermott DS, Alinier G, Charnetski M, Ludlow J, Horsley E, et al. Healthcare simulation standards of best Practice™ simulation design. Clin Simul Nurs. 2021;158:14–21. https://doi.org/10.1016/j.ecns.2021.08.009
28. Decker S, Alinier G, Crawford SB, Gordon RM, Jenkins D, Wilson C. Healthcare simulation standards of best Practice™ the debriefing process. Clin Simul Nurs. 2021;158:27–32. https://doi.org/10.1016/j.ecns.2021.08.011
29. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:1–8.
30. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res. 1992;5(4):194–7.
31. Morita N, Ishihara T, Yamamoto R, Shide N, Okuda T. Content validity and reliability of an enjoyable multicomponent agility test for boys: the N-challenge test. J Sports Sci. 2022;40(9):976–83.
32. Ratter J, Pellekooren S, Wiertsema S, et al. Content validity and measurement properties of the lower extremity functional scale in patients with fractures of the lower extremities: a systematic review. J Patient Rep Outcomes. 2022;6(1):11.
33. Yusoff MSB. ABC of content validation and content validity index calculation. Educ Med J. 2019;11(2):49–54.
34. Gomez R, Brown T, Watson S, Stavropoulos V. Confirmatory factor analysis and exploratory structural equation modeling of the factor structure of the questionnaire of cognitive and affective empathy (QCAE). PLoS ONE. 2022;17(2):e0261914.
35. Della Vedova AM, Loscalzo Y, Giannini M, Matthey S. An exploratory and confirmatory factor analysis study of the EPDS in postnatal Italian-speaking women. J Reprod Infant Psychol. 2022;40(2):168–80.
36. Şencan H. Reliability and validity in social and behavioral measurements. Ankara: Seçkin Publishing. Published online 2005.
37. Esin MN. Veri toplama yöntem ve araçlarıveri toplama araçlarının güvenirlik ve geçerliği. Erdoğan S, Nahçivan N. Esin MN Hemşirelikte Araştırma: Süreç, Uygulama ve Kritik, İstanbul: Nobel Tıp Kitapevi. 2014;217–30.
38. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
39. Meydan CH, Sesen H. Structural equation modeling AMOS applications. Ankara: Detay Publishing. Published online 2011.
40. Lim YJ. Exploratory structural equation modeling analysis of the social network site use motives scale. Psychiatry Investig. 2022;19(2):146.
41. Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs. 2015;18(3):66–7.
42. Bergmann J, Krewer C, Müller F, Jahn K. The scale for retropulsion: internal consistency, reliability and construct validity. Ann Phys Rehabil Med. 2022;65(2):101537.
43. Karagöz Y. SPSS 21.1 Uygulamalı Biyoistatistik. Ankara: Nobel Yayın Dağıtım. 2014;42:203–6.
44. Spence SH, Rapee RM. The development and preliminary validation of a brief scale of emotional distress in young people using combined classical test theory and item response theory approaches: the brief emotional distress scale for youth (BEDSY). J Anxiety Disord. 2022;85:102495.
45. Liu Y, Zumbo BD. The impact of outliers on cronbach's coefficient alpha estimate of reliability: visual analogue scales. Educ Psychol Meas. 2007;67(4):620–34.
46. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. Psychol Assess. 2019;31(12):1412. PMID: 30896212.
47. Chevalier LL, Michaud AL, Zhou ES, Chang G, Recklitis CJ. Validation of the three-item insomnia severity index short form in young adult cancer survivors: comparison with a structured diagnostic interview. J Adolesc Young Adult Oncol. 2022;11(6):596–9. https://doi.org/10.1089/jayao.2021.0175
48. Alpar R. Applied multivariate statistical methods. Ankara: Detay Publishing; 2013.

Publisher's note