Original Article

Psychometric validation of the Creighton Competency Evaluation Instrument in simulation

Validação psicométrica do instrumento *Creighton* para avaliação de competências clínicas em simulação

Validación psicométrica del instrumento Creighton para evaluar competencias clínicas en simulacro

Jéssica de Oliveira Veloso Vilarinho[1] https://orcid.org/0000-0003-0503-7104
Jorge Vinícius Cestari Felix[1] https://orcid.org/0000-0002-0086-674X
Luciana Puchalski Kalinke[1] https://orcid.org/0000-0003-4868-8193
Alessandra Mazzo[2] https://orcid.org/0000-0001-5074-8939
Francisco Diego Negrão Lopes Neto[1] https://orcid.org/0000-0003-2094-5928
Radamés Boostel[1] https://orcid.org/0000-0003-2678-0001
Nilton Orlando da Silva[1] https://orcid.org/0000-0003-1956-7376
Amanda Carolina de Oliveira Bialetzki Fontoura[1] https://orcid.org/0000-0002-4983-7915

How to cite:
Vilarinho JO, Feliz JV, Kalinke LP, Mazzo A, Lopes Neto FD, Boostel R, et al. Psychometric validation of the Creighton Competency Evaluation Instrument in simulation. Acta Paul Enferm. 2020;33:eAPE20200314.

DOI
http://dx.doi.org/10.37689/acta-ape/2020A003146

Keywords
Clinical competence; Psychometrics; Simulation; Validation studies

Descritores
Competência clínica; Psicometria; Simulação; Estudos de validação

Descriptores
Competencia clínica; Psicometría; Simulation; Estudios de validación

Submitted
December 11, 2019

Accepted
April 27, 2020

Corresponding author
Jessica De Oliveira Veloso Vilarinho
Email: jessica.o.veloso@gmail.com

Abstract

Objective: To validate and assess the Creighton Competency Evaluation Instrument reliability, Brazilian Portuguese version.

Methods: This is a methodological study conducted with 30 professors/nurses who used the simulation methodology. Two scenarios were elaborated with equal themes and divergent performances of students in order to test the ability of the instrument to differentiate them. The scenarios were recorded and made available to the experts by a website that enabled the collection. For construct validation, the experimentation and phi correlation techniques were used. For reliability, equivalence with the Fleiss Kappa coefficient and internal consistency with Kuder-Richardson coefficient (KR20) were analyzed.

Results: When analyzing validity, 20 of the 23 items showed significant results. In analysis, the domains presented correlations, classified from slight to substantial. When the scenarios were compared, the correlations were low, poor or negative. Regarding reliability, the Fleiss Kappa coefficient found was fair (0.282) for scenario 1 and moderate for scenario 2 (0.408). The internal consistency measured with KR20 was 0.717 for scenario 1 and 0.805 for scenario 2, classified as moderate and substantial.

Conclusion: The Creighton Competency Evaluation Instrument, Brazilian Portuguese version, was considered valid and reliable for use nationally.

Resumo

Objetivo: Validar construto e avaliar a confiabilidade do Instrumento Creighton para Avaliação de Competências Clínicas, versão português, Brasil.

Métodos: Estudo metodológico, realizado com 30 docentes/enfermeiros que utilizavam a metodologia de simulação. Foram elaborados dois cenários com temas iguais e atuações divergentes das estudantes com o objetivo de testar a capacidade do instrumento de diferenciá-los. Os cenários foram gravados e disponibilizados para os especialistas por um *website* que viabilizou a coleta. Para a validação de construto, utilizou-se as técnicas de experimentação e da correlação phi. Para a confiabilidade, analisou-se a equivalência, com coeficiente Fleiss Kappa, e a consistência interna, com coeficiente Kuder-Richardson (KR20).

Resultados: Ao analisar a validade, 20 dos 23 itens apresentaram resultados significativos. Na análise, os domínios apresentaram correlações, classificadas de baixa a alta. Quando comparados os cenários, as correlações foram baixas, muito baixas ou negativas. Quanto à confiabilidade, o coeficiente Fleiss Kappa encontrado foi razoável (0,282) para o cenário 1 e moderado para o cenário 2 (0,408). A consistência interna medida com o KR20 foi de 0,717 para o cenário 1 e 0,805 para o cenário 2, classificadas como moderada e alta.

[1]Universidade Federal do Paraná, Curitiba, PR, Brazil.
[2]Universidade de São Paulo, Bauru, SP, Brazil.
Conflicts of interest: nothing to declare.

Conclusão: O Instrumento Creighton para Avaliação de Competências Clínicas versão português, Brasil, foi considerado válido e confiável para uso em todo o território nacional.

Resumen

Objetivo: Validar el constructo y analizar la confiabilidad del instrumento Creighton para Evaluación de Competencias Clínicas, versión en portugués, Brasil.

Métodos: Estudio metodológico, realizado con 30 docentes/enfermeros que utilizaban la metodología de simulacro. Se elaboraron dos situaciones con temas iguales y actuaciones divergentes de las estudiantes, con el objetivo de probar la capacidad del instrumento de diferenciarlas. Las situaciones fueron grabadas y se pusieron a disposición de los especialistas mediante un sitio web que viabilizó la recolección. Para la validación del constructo se utilizaron las técnicas de experimentación y del coeficiente de correlación phi. Para la confiabilidad, se analizó la equivalencia con coeficiente Fleiss Kappa y la consistencia interna con coeficiente Kuder-Richardson (KR20).

Resultados: Al analizar la validez, 20 de los 23 ítems presentaron resultados significativos. En el análisis, los dominios presentaron correlaciones, clasificadas de baja a alta. Al comparar las situaciones, las correlaciones fueron bajas, muy bajas o negativas. Respecto a la confiabilidad, el coeficiente Fleiss Kappa encontrado fue razonable (0,282) en la situación 1 y moderado en la situación 2 (0,408). La consistencia interna medida con el KR20 fue de 0,717 en la situación 1 y 0,805 en la situación 2, clasificada como moderada y alta.

Conclusión: El instrumento Creighton para Evaluación de Competencias Clínicas, versión en portugués, Brasil, fue considerado válido y confiable para su uso en todo el territorio nacional.

Introduction

Competency-based training of nurses has been the focus of curricular restructuring since the term incorporation in the 2001 National Curriculum Guidelines for Nursing Graduation.[1,2] From then on, there was a consequent requirement in the inclusion of new pedagogical practices, with inevitable impact on more rigorous professor evaluations in practical activities of nursing students.[2]

In the context of searching for a methodology that allows competency training, clinical simulation emerges as a technique capable of enhancing learning and developing essential clinical competencies.[3,4] Clinical competency is the interconnection of discrete and measurable knowledge, skills and attitudes, essential to the quality of care and patient safety.[5]

Clinical simulation consists of reliable replication of clinical environments, which allows students to develop skills in a safe environment, with self-confidence, satisfaction, and critical reflection.[6] This strategy allows professors to evaluate based on pre-established criteria, in a controlled environment, and that can be repeated as many times as necessary.[7]

With regard to evaluation, the decisions of professors, mainly related to clinical competency evaluation, must be guided by valid and reliable instruments for use in a simulated environment.[8,9] The instruments are essential to identify the knowledge gaps of each student, for a systematized and judicious observation of actions and for neutrality of professors' decisions.[10]

Furthermore, the Creighton Competency Evaluation Instrument (CCEI), which was built with the aim of evaluating clinical competencies in a simulation environment, went through translation and cross-cultural adaptation into Brazilian Portuguese. It was entitled in this version as *Instrumento Creighton para Avaliação de Competências Clínicas* (CCEI-Br), Brazilian Portuguese version.

Competency evaluation in a simulated environment is still a recent practice in Brazil as there is predominance of simulation use for skills assessment, with the use of checklists created by professors, and competencies evaluated in real clinical practice. Moreover, Brazilian research is mostly concerned with proving the efficiency of simulation practice. This translation and adaptation was important, therefore, because until then there were no instruments available in Brazil with this objective and directed to this practice.[11,12]

Thus, this work aims to verify the construct validity and evaluate the reliability properties of CCEI-Br.

Methods

This is a methodological and quantitative study, which followed the Pasquali's methodological framework recommendations.[13] The study was developed in the clinical simulation laboratory of a federal public university in southern Brazil and in a

form online, using a platform for creating free sites, from March to July 2019. The study population consisted of nurses and/or nursing professors from all over the country who used clinical simulation in their practices.

Recruitment took place via e-mail with invitation to experts from the field of clinical simulation in nursing and their indication; through dissemination in research groups and by active search on platform lattes, available at the website of the Brazilian National Council for Scientific and Technological Development (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*, abbreviated CNPQ), using the guide subject and with the descriptors clinical simulation and nursing.

Sample was therefore selected using non-probabilistic sampling for convenience. A total of 221 emails were triggered, with a conversion of 30 professionals who agreed to participate. The instrument used for data collection was the CCEI-Br, a translated and adapted version of the original CCEI, developed in 2008 by the University of Creighton, in the state of Nebraska, U.S.A.[14] Four years after the original version was created, the CCEI was revised to meet the concepts and terminologies of Quality and Safety Education for Nurses (QSEN), with 23 dichotomous classification behaviors between and *demonstrates/does not demonstrates competency* (*demonstra/não demonstra competência*) (0 to 1) and *not applicable* (*não se aplica*). Such classifications were distributed into *evaluation*, *communication*, *clinical judgement,* and *patient safety*.[15]

The CCEI was validated, with 31 teaching experts from the U.S. territory. On this occasion, a Cronbach's alpha of 0.979 and a 79.4% agreement percentage were found among the specialists.[11] In Brazil, in the translation and cross-cultural adaptation stage, a Cronbach's alpha of 0.897 was found, considered an excellent internal consistency and a 100% agreement percentage among specialists.[14]

The operationalization of this study took place in stages, beginning in the elaboration and recording of two clinical simulation scenarios with equal themes and divergent performances. This need arose to respond to the hypothesis created that the instrument would be able to discriminate the divergent actions of two participants, in experimentally produced scenarios, as recommended by Pasquali's experimentation technique for construct validation.

The two recorded scenarios were identical and were based on a simulated practice on altered level of consciousness, caused by hypoglycemia, which contemplated all items of the CCEI and which differed only by the divergent performance of the two students. For the elaboration of the scenarios, Jeffries and Waxman's methodological recommendations were followed.[4,16] The theme chosen followed the latest guidelines of the Brazilian Diabetes Society.[17]

The scenarios were composed of a patient (actor), an intern (student support) and a nurse (student who would be evaluated). When starting the scenario, the student encountered the patient unconscious and the intern in despair because she had just talked to the patient. The objective of the scenario was for the student to identify the cause of the patient's non-responsiveness, identifying vital signs and checking blood glucose, and proceed with the correct intervention, which would be administration of glucose, according to medical prescription and hypoglycemia protocol available.

After audiovisual recording of the scenarios, a pre-test was applied with five invited professors/nurses who knew the simulation method, so that they could evaluate the students' performance in both scenarios. At the time, pre-test participants watched the two recorded scenarios and evaluated the performance of students using the CCEI-Br. Then, with the objective of covering the largest number of participants, a website was built, so that it was possible to provide the participating population with the step by step of how to respond to the survey, the recorded scenarios, the CCEI for evaluating the scenarios and guidance guides.

In data collection, participants who agreed to the Informed Consent Form (ICF) accessed the link on the website available in the invitation email and evaluated the students' performance with the support of the previously prepared discussion sheet. This sheet contained the minimum behaviors expected from each student, for each item, with the objective of establishing an evaluation standard. The collected data was stored in a Microsoft Excel 2016 spreadsheet.

In data analysis, the instrument's construct validity and reliability were evaluated. For construct validity, two techniques were chosen, among the several ways to validate. The first concerns experimentation, described by Pasquali as the best technique for analysis by hypothesis, which is the instrument's ability to discriminate criteria groups. In this case, it was necessary to formulate a hypothesis: the instrument would be able to discriminate the divergent performances of two participants in experimentally-produced scenarios. Thus, comparative statistical analysis was used, with the Statistical Package for the Social Sciences (SPSS®), version 25.0.

The second technique was behavioral analysis of the construct, in which the correlation between items is analyzed, with calculation of the phi correlation coefficient (Φ). This coefficient is the one of choice for dichotomous instruments, ranging from -1 to +1. The interpretation of this coefficient followed the classification: weak (0 <phi <0.1), fair (0.1≤phi <0.3), moderate (0.3≤phi <0.5), strong (0.5≤phi0.7), very strong (0.7≤phi <0.9), and perfect (0.9≤phi≤1).[18] In the statistical analysis, the Free R Statistical Software was used, with the use of the psych package.

In assessing reliability, two criteria were analyzed, according to the framework. The first referred to the test equivalence, which is related to the degree of agreement between observers. In this context, the Fleiss Kappa coefficient of agreement was used, ideal for analyzes with more than two observers, with support from the Free Statistical Software R, irr package. The reference values for this coefficient were poor (<0.00), slight (0.00 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80) and almost perfect (0.81 to 1.00).

The second reliability criterion was internal consistency analysis. To that end, the Kuder-Richardson coefficient (KR20) was used, specific for tests with dichotomous items and which can vary from 0 to 1. The interpretation of the KR20 results followed the slight (KR20 <0.50), moderate (0.50≤KR20 <0.8) and substantial (KR20≥0.80) classification. In this case, the statistical analysis was performed with the aid of the Statistician Free R, DescTools package.

This study met the criteria contained in Resolution 466 of December 12, 2012. The Research Ethics Committee of *Universidade Federal do Paraná* approved the study, under Opinion 3,175,444.

Results

The sociodemographic characterization was presented with a majority of females (80%), with ties to public institutions (76.7%) and with at least a master's degree (90%). The experience of at least two years with the simulation was proven by 76.7% of the participants, of whom 69.6% were more than five years old. Almost all regions had representatives in the sample, with the exception of northern Brazil.

In validation, with analysis by hypothesis, 20 of the 23 items diverged between the two scenarios, with significant results (p <0.05). *Clinical judgement* showed significant results in eight of the nine items, with the exception of item 16 (reflects on clinical experience) (p=0.456). In *patient safety*, two of the six items presented results that reject the elaborated hypothesis (p> 0.05). They are item 20 (*administers medications safely*) and item 23 (*reflects on potential hazards and errors*).

As for behavioral analysis of the construct, there was a correlation between all items, which ranged from moderate to substantial. *Evaluation* showed fair to substantial phi correlation coefficients (phi: 0.370 to 0.557) in both scenario 1 and scenario 2. In *communication*, item 8 (acts professionally) in scenario 2 did not show any correlation, due to invariance in responses. The other items showed a predominance of results considered slight, with phi values between 0.1 and 0.3 for both scenarios.

For *clinical judgment*, the correlations varied from poor to high, with a predominance of slight correlation in scenario 1 and moderate in scenario 2. The results of *patient safety* showed similar correlations to *clinical judgment*. The two scenarios were compared item by item regarding the phi correlation coefficient, as shown in Table 1. Most items showed a poor or negative correlation.

In relation to the reliability test equivalence, the Fleiss Kappa degree of agreement and the percentage of agreement were calculated, by domain and for each scenario, as shown in Table 2 below. Scenarios 1 and 2 showed a Fleiss Kappa level of agreement

Table 1. Correlation matrix of the comparative phi coefficient between items in scenarios 1 and 2

Scenario 1 versus Scenario 2 item2	Phi correlation coefficient	Scenario 1 versus Scenario 2 item	Phi correlation coefficient
1	0.062	13	N/A
2	N/A	14	-0.141
3	0.062	15	0.060
4	0.034	16	-0.449
5	0.064	17	0.208
6	0.204	18	-0.141
7	0.073	19	-0.256
8	N/A	20	-0.254
9	-0.068	21	0.090
10	0.152	22	-0.099
11	-0.082	23	0.156
12	N/A	-	-

N/A - unable to calculate the correlation

of 0.282 and 0.408, considered respectively fair and moderate and the percentage of agreement for scenarios 1 and 2 was 81% and 63%, respectively.

As for the internal consistency to measure the reliability of the CCEI, Table 3 shows that the KR20 reliability coefficient was 0.717 for scenario 1, considered moderate, and 0.805 for scenario 2, classified as substantial consistency. All domains, if excluded, have a negative impact on the general reliability score of the scale.

Patient safety is the one with the greatest variation in KR20, from 0.275 in scenario 1, to 0.625

Table 2. The Fleiss Kappa degree of agreement and percentage of agreement for each scenario

		Total of items	Fleiss kappa	P value	% agreement	Agreement degree
SCENARIO 1	GENERAL	23	0.282	<0.001*	81	Fair
	Evaluation	3	0.022	0.392	89	-
	Communication	5	0.356	<0.001*	88	Fair
	Clinical judgment	9	0.173	<0.001*	86	Mild
	Patient safety	6	0.281	<0.001*	63	Fair
SCENARIO 2	GENERAL	23	0.408	<0.001*	63	Moderate
	Evaluation	3	0.008	0.783	96	-
	Communication	5	0.546	<0.001*	78	Moderate
	Clinical judgment	9	0.225	<0.001*	61	Fair
	Patient safety	6	0.289	<0.001*	36	Fair

p – McNemar test

Table 3. Kuder-Richardson reliability coefficient for each scenario and by domain

	Confidence interval (95%)	Scenario 1 KR20	Scenario 2 KR20
	GENERAL	**0.717**	**0.805**
EVALUATION	ITEMS	0.633	0.494
	1 – Obtains pertinent data	0.434	0.000
	2 – Performs follow-up assessments as needed	0.658	0.659
	3 – Assesses the environment in an orderly manner	0.541	0.000
COMMUNICATION	ITEMS	0.201	0.404
	4 – Communicates effectively with intra/interprofissional team (TeamSTEPPS, SBAR, Written Read Back Over)	0.215	0.340
	5 – Communicates effectively with patient and significant other (verbal, nonverbal, teaching)	0.137	0.187
	6 – Documents clearly, concisely, & accurately	-0.058	0.369
	7 – Responds to abnormal findings appropriately	0.137	0.381
	8 – Promotes professionalism	0.215	0.431
CLINICAL JUDGEMENT	ITEMS	0.601	0.749
	9 – Interprets vital signs (T, PA, FR, FC, Pain)	0.583	0.734
	10 – Interprets lab results	0.609	0.725
	11 – Interprets subjective/objective (recognizes relevant from irrelevant data)	0.510	0.729
	12 – Prioritizes appropriately	0.528	0.761
	13 – Performs evidence based interventions	0.610	0.661
	14 – Performs evidence based rationale for interventions	0.610	0.691
	15 – Evalutes evidence based interventions and outcomes	0.512	0.742
	16 – Reflects on clinical experience	0.497	0.751
	17 – Delegates appropriately	0.627	0.718
PATIENT SAFETY	ITEMS	0.275	0.625
	18 – Uses patient identifiers	0.364	0.636
	19 – Utilizes standardized practices and precautions including hand washing	0.127	0.586
	20 – Administers medications safely	-0.104	0.496
	21 – Manages technology and equipment	0.253	0.598
	22 – Performs procedures correctly	0.475	0.602
	23 – Refletcs on potential hazards and errors	0.069	0.552

in scenario 2. *Clinical judgment*, on the other hand, is the one that presents the best reliability values, if analyzed in isolation, with KR20: 0.601 for scenario 1 and KR20: 0.749 for scenario 2.

Discussion

The results of this study allowed for construct validation and confirmation of the CCEI's reliability. This allows teachers and nurses from all over the country to be provided with an instrument capable of estimating the clinical competence of nursing students and professionals who are evaluated in high-fidelity clinical simulation scenarios. It is believed that these results may contribute especially to the development of clinical competencies while still training nursing students.

With regard to sociodemographic characterization, the sample consisted mostly of female participants, a characteristic closely related to the nursing profession until today, although the proportion of males has increased in classrooms of undergraduate courses.[19] The participating professors and nurses were predominantly from public institutions and had a master's degree as the last degree of instruction. This result is justified by predominance of scientific production in public institutions.[20]

In construct validation, conceptual analyzes of each domain were taken into account. Regarding analysis by hypothesis, *evaluation* and *communication* presented results that confirmed the hypothesis elaborated. These same domains, related to the behavioral analysis of the construct, presented moderate and substantial phi correlation coefficients for *evaluation* and slight for *communication*, in both scenarios. A poor correlation was obtained by comparing the items from the domains in the two scenarios, which allowed us to conclude that the aforementioned domains proved the construct validation.

The phi correlation coefficient is considered the choice for the most current psychometric theory, called Item Response Theory. It is often used in factor analysis, which was not analyzed in this study. For this reason, it is necessary to point out that, in general, samples greater than 50 participants are recommended, which was not the case and which may justify the results below the expected for this coefficient.[21] Moreover, the number of items in each domain directly influences the results of the correlation coefficient.[13] Given this context, it is emphasized that the phi correlation coefficient was not verified in isolation in this study.

With regard to *clinical judgment*, item 16 (reflects on clinical experience) was the only one that presented non-significant results due to the number of *not applicable* options noted. This happened because the evaluators, for mastering simulation, reported in the observation that this item should be analyzed at the time of debriefing. Debriefing is the step that happens after the simulated scenario, which was not covered in the recording.

As for the correlation coefficient of *clinical judgment*, there was a prevalence of slight to moderate correlations in both scenarios, and poor or negative when the scenarios are compared. It is noted that more substantial internal consistencies were highlighted in the results of scenario 2 related to the correlation coefficients, justified by the facilitated judgment of the student's poor performance in this scenario. This resulted in more *does not demonstrate competency* and less *not applicable*.

Regarding the assessment of scenario 1, related to this domain, it was noticed that despite the guidelines to follow the minimum behaviors previously defined, nurses inevitably judged from the perspective of their backgrounds and values and found it difficult to establish a response pattern, as confirmed nursing literature.[22] Based on this, it can be concluded that the category of clinical judgment competence demonstrated findings that allow to confirm construct validation

Regarding *patient safety*, only items 20 (administers medications safely) and 23 (reflect on potential hazards and errors) did not show significant results related to the hypothesis of divergence. Item 23 presents the same justification as item 16 and is related to debriefing. With respect to item 20 (administers medications safely), there were differences of opinion on the performance of the student in scenario 1.

Participants reported in the field observation of the instrument: "I found it difficult to answer some questions as they covered more than one statement and sometimes only one had been followed correctly. For example, she does not wash her hands, but uses the glove" (participant 13). The patient safety topic consists of a complex discussion involving one of the international goals for ensuring patient safety, established by the World Health Organization and therefore subject to greater rigor in the performance evaluation.[23]

In relation to the correlation coefficient of this domain, substantial coefficients were found in scenario 2 as well as in the other domains and prevalence of slight to moderate correlations. When comparing the scenarios, the items in this domain showed poor or negative correlations, which allowed us to conclude that *patient safety* presented results that confirm the construct validation.

The CCEI reliability, related to equivalence, was confirmed with the Fleiss Kappa coefficient of agreement considered fair for scenario 1 and moderate for scenario 2, with values of 0.282 and 0.408. In validating the original scale, the authors found agreement coefficients also classified as fair and moderate, with values 0.316, 0.453 and 0.443.[15]

The CCEI validated in the Spanish version found Kappa coefficient of agreement values of 0.80, but it is noteworthy that the study finalized the scale with only 22 items and the choice analysis method was different from the current study.[24] The Fleiss Kappa coefficients of the domains in this study showed results inversely proportional to the agreement percentages; were justified by analysis of existence of random error that the Fleiss Kappa coefficient uses, in contrast to the percentage of agreement, which only assesses the gross agreement of the evaluators.[25]

Internal consistency, calculated with the KR20 reliability coefficient, showed values of KR20: 0.717 for scenario 1 and KR20: 0.805 for scenario 2, was classified as moderate and substantial, respectively. The original version of the CCEI found internal consistency values greater than 0.90. It is noteworthy that the coefficient used in this case was Cronbach's Alpha.[15]

In the Spanish version, the internal consistency found was 0.839, classified as substantial.[24] About the Brazilian version, still in the translation stage, Cronbach's alpha of 0.897 was found.[14] Some authors found lower values of internal consistency when using the KR20 coefficient, with values of 0.701 and 0.803, for dichotomous scales, considered moderate and substantial consistency.[26]

The validation and reliability verification process followed the guidelines of the methodological framework and presented limitations with regard to the availability of other valid and reliable instruments. Such instruments were built with the same objective, and that would allow verification of other types of validation. Furthermore, the study had limitations regarding the recruitment of participants, which, due to being via email, ended up allowing a slight conversion of those who agreed to contribute to the research.

Conclusion

This study met the objective of validating the construct and testing the reliability of the Brazilian Portuguese version of the CCEI. The statistical and theoretical analyzes carried out prove that the instrument was able to discriminate the divergent scenarios. In addition, the findings of the correlation coefficients were considered acceptable, in association with the theoretical discussion. These results confirmed the instrument's construct validation. Reliability analysis showed equivalence parameters considered, in general, fair and moderate, for the two scenarios and parameters of internal consistency considered moderate and substantial. Thus, it can be concluded that the referred instrument has good internal consistency and is considered reliable to measure the clinical competence of nursing students/professionals, when evaluated in clinical simulation.

Collaborations

Vilarinho JOV, Felix JVC, Kalinke LP, Mazzo A, Lopes Neto FDN, Boostel R, Silva NO and

Fontoura ACOB declare that they contributed to the project design, analysis and interpretation of data, writing of the article, relevant critical review of the intellectual content and approval of the final version to be published.

References

1. Brasil. Conselho Nacional da Educação. Câmara de Educação Superior. Resolução CNE/CES 3/2001. Brasília (DF): Ministério da Saúde; 2001. [citado 2019 nov 01]. Disponível em: http://portal.mec.gov.br/cne/arquivos/pdf/CES03.pdf
2. Winters JR, Padro ML, Lazzari DD, Jardim VL. A formação superior em enfermagem no MERCOSUL: um estudo bibliométrico [Internet]. Rev Bras Enferm. 2018;71 Suppl 4:1732–9.
3. Rosa RS, Sanches GJC, Gomes ICR, Silva ML, Duarte ACS, Boery RNSO. Estratégias baseadas em metodologias ativas no ensino-aprendizagem de primeiros socorros: relato de experiência. Rev Enferm UFPE Online. 2017;11(2):798-803.
4. Jeffries PR, Rodgers B, Adamson K. NLN Jeffries Simulation Theory: brief narrative description. Nurs Educ Perspect. 2015 ;36(5):292–3.
5. Meakim C, Boese T, Decker S, Franklin AE, Gloe D, Lioce L, et al. Standards of best practice: simulation. Clin Simul Nurs. 2013;9(6):3–11.
6. Costa RRO, Medeiros SM, Martins JCA, Enders BC, Lira ALBC, Araújo MS. A simulação no ensino de enfermagem: uma análise conceitual. Rev Enferm Centro-Oeste Mineiro. 2018;8(e1928):e19288.
7. Escudero E, Bem-Azul MA, Cancino KD. Simulación clínica y seguridade del paciente: integración em el currículo de enfermería. Sci Med (Phila). 2018;28(1):1–13.
8. Bergamasco EC, Murakami BM, Cruz DA. Uso da Escala de Satisfação dos Estudantes e Autoconfiança com a Aprendizagem (ESEAA) e da Escala do *Design* da Simulação (EDS) no ensino de enfermagem: relato de experiência. Sci Med (Phila). 2018;28(3):1–5.
9. Presado MH, Colaço S, Rafael H, Baixinho CL, Félix I, Saraiva C, et al. Aprender com a simulação de alta fidelidade. Cien Saude Colet. 2018;23(1):51–9.
10. Gomes JÁ, Martins MM, Tronchin DM, Fernandes CS. Validação de uma escala sobre indicadores de estrutura no bloco operatório: contributos para a enfermagem. Arquichan. 2019;19(1):1–13.
11. Miranda FBG, Mazzo A, Pereira Junior GA. Avaliação de competências individuais e interprofissionais de saúde em atividades clínicas simuladas: scoping review. Interface comum. Saúde Educ. 2018;22(67):1221-34.
12. Sebold LF, Boell JEW, Girondi JBR, Santos JLG. Simulação clínica: desenvolvimento de competência relacional e habilidade prática em fundamentos de enfermagem. Rev Enferm UFPE Online. 2017;11(Supl 10):4184-90.
13. Pasquali L. Validade dos testes. Revista Examen. 2017;1(1):14–48.
14. Silva NO. Tradução e adaptação transcultural do Creighton Competency Evaluation Instrument para a língua portuguesa (Brasil) [dissertação]. Curitiba: Universidade Federal do Paraná; 2019.
15. Hayden J, Keegan M, Kardong-Edgren S, Smiley RA. Reliability and validity testing of the Creighton Competency Evaluation Instrument for use in the NCSBN National Simulation Study. Nurs Educ Perspect. 2014;35(4):244–52.
16. Waxman KT. The development of evidence-based clinical simulation scenarios: guidelines for nurse educators. J Nurs Educ. 2010;49(1):29–35.
17. Oliveira JP, Motenegro RM Junior, Vencio S. Diretrizes da Sociedade Brasileira de Diabetes 2017-2018. São Paulo: Editora Clannad; 2017.
18. Burns MK, Zaslofsky AF, Kanive R, Parker DC. Mata-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. J Behav Educ. 2012;21(1):158–202.
19. Correa Júnior AJ, Santana ME, Farias GM, Noronha EC. Globalizacíon em la postmodernidad: críticas y contribuciones para la enfermería. Cultura de los Cuidados. 2019;23(53):265-74.
20. Smolski FMS, Dalcin D, Visentini MS, Bamberg J, Kern JS. Análise do perfil da produção científica da Revista de Administração Pública (RAP). Rev. adm. pública. 2020. [cited 2020 abr. 15]. Available from: http://www.scielo.br/pdf/rap/v51n6/1982-3134-rap-51-06-1139.pdf
21. Collares CF, Grec WL, Machado JL. Psicometria na garantia de qualidade da educação médica: conceitos e aplicações. Science in Health. 2012;3(1):33–49.
22. Santos VR. Avaliação escolar – uma prática entre duas lógicas: a que emancipa e a que reproduz as desigualdades sociais. Interação (Curitiba). 2019;44(2):515–28.
23. Cardoso AS, Muller S, Echer IC, Rabelo-Silva ER, Boni FG, Ribeiro AS. Elaboração e validação do *checklist* para administração de medicamentos para pacientes em protocolo de pesquisa. Rev Gaúcha Enferm. 2019;40(spe):1-5.
24. Roldán-Merino J, Farrés-Tarafa M, Estrada-Masllorens JM, Hurtado-Pardos B, Miguel-Ruiz D, Nebot-Bergua C, et al. Reliability and validity study of the Spanish adaptation of the "Creighton Simulation Evaluation Instrument (C-SEI)". Nurse Educ Pract. 2019;35(1):14–20.
25. Kuo BI. Intraclass correlation coefficient rather than correlation coefficient to examine agreements among different methods measuring valvular area. Circulation. 1994;89(4):1910–1.
26. Vizzotto PA, Mackedanz LF. Validação de isntrumento de avaliação da alfabetização científica para egressos no ensino médio no contexto da física do trânsito. Educ Rev. 2018;34:e202974.