Validation of the Dutch version of the Nurses Clinical Reasoning Scale to evaluate nurses' perception of clinical reasoning competence

Name of student: Janssen, B
Student number: 6508790
Status research proposal: final version
Date: 21/06/2021
University of Utrecht, Master's program KGW, Nursing Science, UMC-Utrecht
Name of teacher: Dr. Janneke de Man-van Ginkel
Name of supervisor: Marloes Veenstra, MSc
Internship institution: Julius Center for Health Sciences and Primary Care, Utrecht
Magazine: International Journal of Nursing studies
Number of words (Chapters Safety Reporting and Administrative aspects, reference list, summary and appendices excluded): 3232
Number of words abstract: 300
Number of words Dutch abstract: 299
Reference style: Vancouver
Criteria used for transparent reporting: COSMIN

ABSTRACT

Title: Validation of the Dutch version of the Nurses Clinical Reasoning Scale to evaluate nurses' perception of clinical reasoning competence.

Background: Clinical reasoning is the fundamental base of clinical practice. It guides nurses in gathering, assessing, interpreting information, and generating hypotheses. Clinical reasoning is of significant importance for patient safety and is an essential element of competence. The Nurses Clinical Reasoning Scale (NCRS) was developed as an evaluation tool for the self-assessment of clinical reasoning competence. To date, the validity and reliability of the translated Dutch version of the NCRS (D-CRS) has yet to be determined.

Aim/RQ: This study aims to test the validity and reliability of the D-CRS for clinical nurses working in the Netherlands.

Method: This study had a quantitative, descriptive, retrospective cross-sectional research design. A secondary analysis using consisting data of a sample of clinical nurses working in a University Medical Center and a general hospital (N= 515) was conducted. Confirmatory factor analysis (CFA) and exploratory factor analysis (EFA) were used to assess construct validity. Reliability was assessed with Cronbach's alpha.

Results: CFA did not provide the hypothesized one-factor structure. EFA showed a two-factor structure. A one-factor model based on 2 factors with high factor correlation (>0.70) was shown to be valid based on Comparative Fit Index (*CFI*=0.95), Tucker-Lewis Index (*TLI*=0.94) and reliable based on Cronbach's alpha=0.94. However, Root Mean Square Error of Approximation (*RMSEA*) did not indicate a good model fit.

Conclusion: The D-CRS showed good reliability and acceptable validity based on CFI/TLI when it is used as a scale to assess clinical reasoning skills in Dutch clinical nurses working in a general and University Medical Center in the Netherlands.

Recommendations: Although the D-CRS should be investigated further in future research, it could be a useful tool for Dutch nurses and their health care managers to assess and improve nurses' clinical reasoning skills in clinical practice.

Keywords: Clinical reasoning competence, nurses, patient safety, reliability, validity.

SAMENVATTING

Titel: Validatie van de Nederlandse versie van de *Nurses Clinical Reasoning Scale* om de perceptie van klinische redeneervaardigheden van verpleegkundigen te evalueren.

Achtergrond: Klinisch redeneren is belangrijk voor verpleegkundigen in de dagelijkse praktijk. Het helpt verpleegkundigen om patiëntgegevens te verzamelen, beoordelen en op basis hiervan te handelen. Klinisch redeneren is ook van belang voor de patiëntveiligheid. De *Nurses Clinical Reasoning Scale (NCRS)* is een zelf-beoordelingsschaal die is ontwikkeld voor verpleegkundigen om hun klinische redeneervaardigheden te beoordelen. Tot op heden is de betrouwbaarheid en validiteit van de Nederlandse versie van de NCRS (D-CRS) nog niet getest in de klinische praktijk.

Doel: Deze studie test de betrouwbaarheid en de validiteit van de D-CRS voor verpleegkundigen die werken in algemeen en academisch ziekenhuis in Nederland.

Methode: Deze studie had een kwantitatief, beschrijvend, retrospectief cross-sectioneel design. Er werd een analyse uitgevoerd met bestaande gegevens van verpleegkundigen werkzaam in een academisch en algemeen ziekenhuis in Nederland (N=515). Bevestigende en verkennende factoranalyse werden uitgevoerd om constructvaliditeit te beoordelen. De betrouwbaarheid werd beoordeeld met Cronbach's alpha.

Resultaten: Bevestigende factoranalyse vertoonde niet de één-factor structuur zoals voorspeld. Exploratieve factoranalyse vertoonde een twee-factor structuur. Een één-factormodel gebaseerd op 2 factoren met hoge factorcorrelatie (> 0.70) bleek valide te zijn op basis van *Comparative Fit Index* (*CFI* = 0.95), *Tucker-Lewis Index* (*TLI* = 0.94) en betrouwbaar op basis van Cronbach's alpha = 0.94. Echter, liet de *Root Mean Square Error of Approximation (RMSEA*) geen goede fit van het model zien.

Conclusie: De D-CRS heeft een goede betrouwbaarheid en acceptabele validiteit op basis van CFI en TLI wanneer de schaal wordt gebruikt om klinische redeneervaardigheden te beoordelen bij verpleegkundigen werkzaam in een algemeen en academisch ziekenhuis in Nederland.

Aanbevelingen: Alhoewel verder onderzoek nodig is naar de D-CRS, kan het voor Nederlandse verpleegkundigen en hun leidinggevende een nuttig hulpmiddel zijn om klinische redeneervaardigheden van verpleegkundigen werkzaam in de klinische praktijk te beoordelen en te verbeteren.

Trefwoorden: Klinisch redeneervaardigheden, verpleegkundigen, patiëntveiligheid, betrouwbaarheid, validiteit.

INTRODUCTION AND RATIONALE

Clinical reasoning is the foundation of clinical practice[1] and important to nursing professionals as it is a problem-solving process that guides nurses in gathering, assessing, interpreting information, and generating hypotheses[2-4]. Nurses play an important role in early detection of complications and impending patient deterioration as they are in the best position to initiate actions to prevent adverse patient outcomes[4]. Clinical reasoning by nurses is key to effective surveillance and may possibly explain the link between higher nursing skills and better patient outcomes[4].

Clinical reasoning can be defined as *"a complex cognitive process that uses formal and informal thinking strategies to gather and analyze patient information, evaluate significance of this information and weigh alternative actions*[2,5]". It is a fundamental feature of health care practice and can be viewed as the hallmark of the expert nurse[6].

Clinical reasoning is an essential skill for nurses to provide safe, effective and quality patient care[2,7]. Previous research shows a difference between nurses with effective clinical reasoning skills and those with poor clinical reasoning skills when it comes to anticipating and identifying patients' deterioration[3-4,8]. Poor clinical reasoning skills can be associated with serious adverse events and failure-to-rescue[3-4,9]. Furthermore, incorrect reasoning can contribute to patient morbidity and mortality[10]. This is supported by the New South Wales (NSW) Health incident Management System, stating that the top three reasons for adverse patient outcomes are: failure to properly diagnose, failure to institute appropriate treatment, and inappropriate management of complications[9]. Therefore, clinical reasoning has significant implications for patient safety[10] and is an essential component of competence[3,6,11].

Despite the fact that clinical reasoning is considered an important component of nursing competence[3,6,11], there is a lack of valid and reliable evaluation tools for the self-assessment of clinical reasoning competence[2]. These insights have recently led to the development of the Nurses Clinical Reasoning Scale (NCRS)[2]. Such an evaluation tool can help nurses identify personal areas for improvement and can also assist healthcare managers to identify nurses' needs and help them become more competent in clinical reasoning, leading to safe and quality patient care[2,7]. The NCRS was developed and psychometrically tested in Chinese and translated into English. The results of this psychometric study supported the validity and reliability of the developed scale[2]. Construct validity of the NCRS was supported based on factor analysis where the emerged one factor explained 50.7% of the variance of clinical reasoning competence. The Cronbach's alpha for the NCRS was good ($\alpha = 0.90$) and the test-retest reliability was evidenced by high intraclass correlation (ICC = 0.87)[2]. Thereafter, it

was concluded that the NCRS is a useful tool and can easily be administered for the self-assessment of clinical reasoning competence of clinical nurses[2]. Due to the described properties, the NCRS seemed well-suited to assess clinical reasoning competence in Dutch clinical nurses. The validity and reliability of a Dutch version of the NCRS (D-CRS) have not yet been determined. The D-CRS could become a useful tool for both Dutch nurses and their healthcare managers to assess and improve nurses' clinical reasoning skills in order to provide safe and good quality patient care. Therefore, this study aims to test the reliability and validity of the D-CRS.

AIM

The aim of this study is to test the validity and reliability of the Dutch version of the Nurses Clinical Reasoning Scale designed for the self-assessment of clinical reasoning competence of clinical nurses working in the Netherlands.

METHOD

Study design and participants

This study had a quantitative, descriptive, retrospective, cross-sectional research design. The D-CRS was validated by testing its construct validity and internal consistency by performing a secondary analysis using existing data. The sample of this dataset consisted of clinical nurses working at the various included nursing departments of a University Medical Center and a general hospital in the Netherlands. All Dutch speaking and registered clinical nurses aged 18 years and older and with educational level 4 or higher[12] were included. Educational level was based on the Dutch National Qualifications Framework (NLQF). The NLQF consists of eight levels and one entry level. The levels are based on descriptions of what someone knows and is able to do after completion of a learning process[12]. NLQF level 4 describes higher general secondary education and higher general secondary education for adults (Dutch: middelbaar beroepsonderwijs niveau 4). Higher educational levels describe pre-university educations, associate degree, bachelor's degree, master's degree and doctorate/medical specialist[12]. This study was conducted in accordance with the COSMIN guidelines[13].

Procedures and data collection

Data for validating the D-CRS was collected from mid-2018 to the end of 2019. In the context of a research into job differentiation for nurses, multiple questionnaires including the D-CRS were administered within various nursing departments at University Medical Centres and general hospitals in the Netherlands[14]. After data was collected, it was stored in a SPSS file, which was made available and used for the validation purposes of this current study. The dataset used for the secondary analysis of this current study was chosen based on sample size requirements for testing construct validity and internal consistency[13,15] and equal ratio between nurses working within a general hospital and University Medical Center. According to these sample size requirements a minimum of 150 participants was recommended[13,15].

Measures

The D-CRS is a 15-item self-report measure that was originally developed for self-assessing clinical reasoning skills in clinical nurses. The items of the scale can be rated using a five-point Likert scale ranging from 1 "strongly disagree" to 5 "strongly agree". A higher score indicates a higher level of clinical reasoning competence[2]. In order to be able to use the questionnaire in Dutch health care settings, it has been translated into Dutch (appendix 1). The D-CRS was translated through the standard approach of the "forward" and "backward" translation which was in accordance with the COSMIN-guidelines[13,15]. Translation was performed by a certified translation agency (VU Amsterdam translation center).

Construct validity and internal consistency

Due to a lacking gold standard, construct validity was measured in order to provide evidence of validity[15]. The degree to which the scores of the D-CRS were consistent with hypothesis was tested by two aspects of construct validity: structural validity and cross-cultural validity[15]. Cronbach's alpha was used to assess the internal consistency of the D-CRS to reveal the degree to which the items are correlated to measure the same construct[16].

Analysis

Descriptive statistics were calculated for participants demographic characteristics. Means and standard deviation were used for continuous variables, and frequency and percentages were used for categorical data. Statistical analysis included Cronbach's alpha to examine the reliability of the D-CRS and a confirmatory factor analysis (CFA) to investigate construct validity. For this study it was hypothesized that the one-factor structure of Liou and colleagues[2] would be replicated. Initially the 15-item D-CRS was tested through CFA. However, for some measurement properties additional criteria can also be appropriate. If CFA shows inadequate model fit an exploratory factor analysis (EFA) is appropriate to assess the results of a study[13]. Therefore, after CFA an EFA was performed in order to determine whether alternative models might equally well or even better fit the data.
CFA was performed in order to test the priori hypothesis about the factor structure and subsequently to validate the resulting dimensions from the EFA. The number of factors to retain from EFA was determined using the parallel analysis (scree-plot) and the eigenvalue-greater-then-one decision rule[15] with direct oblimin rotation. The use of CFA to test construct validity adds a level of statistical precision[13]. Maximum Likelihood Estimation was used to assess model fit. A full dataset was used with no missing values. To assess the strength of the CFA model, the adequacy of model fit was determined based on three fit indices: 1) Comparative Fit Index (CFI), 2) Tucker Lewis Index (TLI) and 3) Root Mean Square Error of Approximation (RMSEA)[17-20]. The following cut-off values were representative of good-fitting models: CFI and TLI higher than 0.90 indicate acceptable fit, whereas values close to 0.95 or higher indicate good fit[15,19-23] and RMSEA between 0.05 and 0.08, with a value of 0.06 or less was considered as an excellent fit[17,19]. Internal consistency was regarded as sufficient if the value of Cronbach's alpha was between 0.70 and 0.95[15-16,24]. The level of significance was specified at 0.05. Data were analyzed using SPSS 26 and Mplus.

Ethical issues

This study was conducted according to the principles of the Declaration of Helsinki[25] (64th version, 2013) and was in accordance with the EU General Data Protection Regulation (GDPR) and other guidelines and regulations. The study protocol was classified by the

quality coordinator of the University Medical Center of Utrecht (UMCU) as “not medical scientific” research and therefore ethical approval was deemed not to be necessary. For the study on job differentiation for nurses[14], the participants gave permission to use the data that emerged from the study for possible follow-up research in the context of the research program RN2BLEND.

RESULTS

Participants

The sample used to perform factor analysis consisted of 515 registered clinical nurses working within various departments in a University Medical Center and a general hospital in the Netherlands. The sample was evenly distributed with 302 nurses working in a general hospital and 213 nurses working in a University Medical Center. This was an adequate sample size as it met the requirements for performing a factor analysis as well as determining internal consistency[13,15]. The mean age of the participants was 40.6 years old (*SD* = 12.3). The clinical nurses had on average 17.9 years of work experience (*SD* = 12.8). Key characteristics of the participants are presented in table 1.

Insert table 1.

Findings

Overall, means per item were high. All but one item had an average score of 4 and higher. The overall Cronbach's alpha was 0.94. CFA examined three models to evaluate the best fit for the overall data (table 2). Model 1 consisted of all items and was used to test if the hypothesized one-factor structure[2] was replicated. Model 1 showed the following values: RMSEA = 0.19, CFI = 0.92 and TLI = 0.91. As these results did not meet all cut-off criteria, an EFA was performed to assess factor structure and dimensionality. Results from the EFA showed two factors based on eigenvalue > 1 and the scree-plot: *1) 'collecting patient data'* and *2)* '*acting based on collected patient data'*. A total of 60.4% of the variance was explained by the two factors. Item loadings were overall high (>0.55) and loaded on their respective factors. The items loaded on *'collecting patient data'* ranged between 0.63 and 0.80 (1 t/m 8). The items loaded on *'acting based on collected patient data'* ranged between 0.56 and 0.80 (10 t/m 15). Some items loaded on the same factors (item 5 t/m 8). Item 9 loaded just below 0.4 (0.32). A minimum loading of 0.5 is usually taken as a threshold[15]. However, item 9 suited best with factor 2 and was therefore not deleted. The two-factor structure derived from EFA was tested for model fit by CFA in Mplus. The results showed that the two-factor models (model 2 and 3) had the most acceptable fit with CFI = 0.95 and TLI = 0.94. However, the RMSEA value was not between 0.05 and 0.08 and therefore not representative of a good-fitting model. Model 3 consisted of the same two factors as model 2 and showed the same results. However, in this model the factor correlation between the two factors was considered. Due to high factor correlation between the two factors, it was considered as a one-factor model, as a high correlation means the two factors measure the same construct (see table 3). Therefore, model 3 was more in line with the original and hypothesized one-factor structure[2]. The assumption of normality was fulfilled.

Insert table 2 and 3.

DISCUSSION

To the best of our knowledge, this is the first study investigating the validity and reliability of the D-CRS. The current study showed good overall psychometric properties of this scale. First, results showed that the D-CRS has good internal consistency. Second, this study showed that the hypothesized one-factor structure of Liou et al[2] had no excellent fit based on all fit indices (CFI, TLI, RMSEA). Further explorative analysis revealed a two-factor structure with a better model fit than the original one-factor structure based on fit indices; CFI and TLI. However, these factors were highly correlated which meant the two factors measured the same construct. A one-factor model based on the aforementioned two factors showed the best model fit, however the RMSEA value remained high.

The observed good internal consistency with a Cronbach's alpha of 0.94 showed that the DCR-S has good reliability. This Cronbach's alpha value may be explained by the fact that the items in the scale are highly correlated[15]. This indicates that the items of the scale measure one unidimensional construct. This is in line with the results from the original study were the NCRS was developed and psychometrically tested, and Cronbach's alpha was 0.90[2]. Additionally, a study conducted in Italy, where the NRCS was linguistic validated and culturally adapted showed the same results with a Cronbach's alpha of 0.90. The researchers concluded that the Italian version can be proposed as an interesting means of evaluating nursing students and nurses in their daily clinical practice[26].

Results showed mixed (adequate and inadequate) fit indices for all three tested models. Despite the fact that CFI and TLI showed acceptable fit (>0.90), the RMSEA was high (>0.10) and therefore did not meet adequate model fit[15,19-23]. As literature describes it is important to not automatically disregard the model just because one index fails to meet the cutoff, nor should researchers retain the model by reporting only the adequate fit indices. Instead, researchers must explain why the fit indices disagree and the implications of the disagreement[27]. A possible explanation for less adequate model fit found in the current study could be differences in culture. The original instrument was tested in Taiwan. However, in the current study the D-CRS was implemented in the Netherlands. The study of Gjersing and colleagues[28] showed that cultural differences and/or the translation of a scale are important factors that could contribute to less adequate model fit[28]. This could also suggest that the cross-cultural generalizability of the D-CRS is limited. The high RMSEA might be due to the fact that cutoff values for RMSEA are developed for continues data and are less suitable for categorical data[29-30]. Similarly, there are some suggestions that CFI and TLI may be better at identifying model fit without over factoring and are as reliable as chi-square test and RMSEA[29]. Data used for the current study are categorical. This may be a reason for the high

RMSEA and less acceptable model fit[29]. Another possible explanation for the high RMSEA can be a small model size (i.e., the total number of items of a scale). According to Shi and colleagues the RMSEA is sensitive to model size[31]. The D-CRS is a small model which consists of 15 items. Therefore, the high RMSEA should be cautiously interpreted[31]. Given the fact that this is the first study testing these psychometric properties of a translated version of the NCRS, we are not able to compare our results to equal studies. However, when comparing the three tested models with CFA, model 3 showed best model fit with higher CFI and TLI and lower RMSEA. Considering more than a one-factor model may be meaningful in future studies for this population to improve model fit.

Strengths and limitations

The findings of the current study should be interpreted with caution as a result of the following limitations. First, the results of the current study might be due to self-assessment bias, an overly positive assessment of personal performance[32]. Almost all items of the D-CRS had a mean of four and higher. The presence of these biases may depend on the type of survey. The D-CRS is a self-report measure. It is plausible that nurses are not likely to say they are bad at clinical reasoning. Therefore, the use of a self-assessment scale may be a confounding factor in the current study[33]. In addition, Liou and colleagues also described that the interpretation of clinical reasoning competences measured by the D-CRS is limited as the scale measures self-perceived competences rather than evidenced based practical competences[2]. Therefore, it is suggested to cautiously interpret clinical reasoning competence measured by the D-CRS. However, for many variables, such as perceptions, motivations and attitudes, self-report is the only direct way to obtain information[34]. A solution to reduce self-assessment bias in future research could be to have an additional assessment carried out by a healthcare manager and/or colleague. Another limitation is the use of the English version of the NCRS for the translation and validation of the D-CRS. The original Chinese version of the NCRS was developed and psychometrically tested, whereas the English version of the NCRS is not. The original Chinese version of the NCRS was translated by Liou and colleagues into English for publication purposes[2]. Therefore, Liou and colleagues[2] also stated that future researchers if plan to use the English version of the NCRS for research purpose need to carefully test its psychometric properties using their target population[2]. Which was done by the researchers of the current study. Lastly, a limitation of the current study is the limited generalizability of the findings. The current study involved only nurses of two hospitals in the Netherlands. Although the sample size was adequate[13,15], the participants were less diverse as there were little recovery- and emergency room (ER) nurses as opposed to general and intensive care unit (ICU) nurses (see table 1). Therefore, it is recommended to use more generalizable data in follow-up research.

A strength is the design of the current study and therefore the use of classical test theory[13]. The design of the current study is in line with COSMIN guidelines[13,15]. COSMIN guidelines and checklists were developed by an international Delphi study[13]. A second strength is the additional exploratory factor analysis, which revealed another factor structure than the original Chinese version of the NCRS. This is important for further research and implementing this scale in other countries. Another strength of the current study is the very large sample size of 515 participants. Especially, considering the minimal number of 150 participants according to sample size requirements[13,15]. Lastly, the current study contributed to the development of a reliable measure of clinical reasoning competence of Dutch clinical nurses working in a general hospital and University Medical Center in the Netherlands.

Conclusion and implications for clinical practice for further research

The D-CRS showed good reliability and acceptable validity when it is used as a scale to assess clinical reasoning competence in Dutch clinical nurses working in a general hospital and University Medical Center in the Netherlands. CFA showed a satisfactory fit based on CFI and TLI, but an unsatisfactory fit based on the RMSEA. Therefore, before implementing the D-CRS it is required to assess model modification to obtain a better-fitting model. This is due to the fact that the validity of the scale cannot be guaranteed based on the current analysis and outcomes. However, the D-CRS is a short and easy administrable questionnaire with good overall psychometric properties. Therefore, the D-CRS could be a useful tool for Dutch clinical nurses and their healthcare managers to assess and improve nurses' clinical reasoning skills in clinical practice. Future research could investigate the possibility to implement an additional assessment by a healthcare manager to assess the clinical reasoning competence of clinical nurses.

REFERENCES

(1) Higgs J, Jones MA, Loftus S, Christensen N. 2008. Clinical reasoning in the health professions. 3rd ed. Edinburgh: Elsevier.

(2) Liou S.-R, Liu H.-C, Tsai H.-M, Tsai Y.-H, Lin Y.-C, Chang C.-H, et al. The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. Journal of Advanced Nursing (JAN) 2015 August 12;707-117.

(3) Levett-Jones T, Hoffman K, Dempsey J, Yeun-Sim Jeong S, Noble D, Norton CA, et al. The 'five rights' of clinical reasoning: An educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. Nurse education today 2009 Oct 30:515-520.

(4) Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality. Journal of the American Medical Association (JAMA) 2003 Sep 24;290(12):1617-1623.

(5) Simmons B. Clinical reasoning: concept analysis. Journal of Advanced Nursing (JAN) 2010;66(5):1151-1158.

(6) Banning M. Clinical reasoning and its application to nursing: Concepts and research studies. Nurse education in practice. 2008;8(3):177-183.

(7) de Carvalho EC, Oliveira-Kumakura ARS, Morais SCRV. Clinical reasoning in nursing: teaching strategies and assessment tools. Revista Brasileira de Enfermagem. 2017 Jun;70(3):662-668.

(8) Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebø Å. Teaching clinical reasoning and decision-making skills to nursing students: Design, development, and usability evaluation of a serious game. International journal of medical informatics (Shannon, Ireland) 2016;94:39-48.

(9) NSW Health. NSW Patient Safety and Clinical Quality Program. Journal of the Australasian Association for Quality in Health Care. 2005 July 26;19(2):4.

(10) Amey L, Donald KJ, Teodorczuk A. Teaching clinical reasoning to medical students. Britisch Journal of Hospital Medicine. 2017 Jul 2;78(7):399-401.

(11) Thampy H, Willert E, Ramani S. Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid. J Gen Intern Med. 2019 Apr 25;34(8):1631-1636.

(12) The Dutch Qualifications Framework (NLQF). Classification of Dutch qualifications in the NLQF and EQF. Available from: https://www.nlqf.nl/nlqf-niveaus. [accessed 5th November 2020].

(13) Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, de Vet H, et al. COSMIN Study Design checklist for patient-reported outcome measurement instruments. July 2019.

(14) RN2BLEND. Registered nurses to Blend [internet]; Available from: https://rn2blend.nl/nl/aan-de-slag. [accessed 28th October 2020].

(15) de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. 10th ed. Cambridge: Cambridge University press; 2018.

(16) Terwee CB, Bot SDM, de Boer MR, van der Windt, Daniëlle A. W. M, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology 2007;60(1):34-42.

(17) Raniti MB, Waloszek JM, Schwartz O, Allen NB, Trinder J. Factor structure and psychometric properties of the Pittsburgh Sleep Quality Index in community-based adolescents. Sleep Research Society (New York, N.Y.) 2018;41(6).

(18) Ballangrud R, Husebo SE, Hall-Lord ML. Cross-cultural validation and psychometric testing of the Norwegian version of the TeamSTEPPS (R) teamwork perceptions questionnaire. BMC Health Services Research 2017; 17:1.

(19) Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999;6(1):1–55.

(20) Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of educational research (Washington, D.C.) 2006;99(6):323-338.

(21) Vedsted P, Sokolowski I, Heje HN. Data quality and confirmatory factor analysis of the Danish EUROPEP questionnaire on patient evaluation of general practice. Scandinavian journal of primary health care 2008;26(3):174-180.

(22) Thombs BD, Hudson M, Schieir O, Taillefer SS, Baron M. Reliability and validity of the center for epidemiologic studies depression scale in patients with systemic sclerosis. Arthritis and rheumatism 2008 Mar 15;59(3):438-443.

(23) Brown GTL, Harris LR, O'Quin C, Lane KE. Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. International journal of research & method in education. 2017 Jan 1;40(1):66-90.

(24) Sun Y, Liu Y, Su T, Yuan J, Liu Z. Medical, epidemiologic, and social aspects of aging urinary incontinence questionnaire. Medicine; 2019 September 28: 98:44

(25) The World Medical Association, (WMA). Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. Journal of international biotechnology law 2013;6(2):73-76.

(26) Notarnicola I, Rocco G, Pulimeno A, De Jesus Barbosa, M R, Lacorossi L, Petrizzo A, et al. Linguistic validation and cultural adaptation of the Nurse Clinical Reasoning Scale. Professioni Infermieristiche 2020 Jan-Mar;73(1):5-11.

(27) Lai K, Green SB. The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree. Multivariate behavioral research 2016 May 03;51(2-3):220-239.

(28) Gjersing L, Caplehorn JRM, Clausen T. Cross-cultural adaptation of research instruments language, setting, time and statistical considerations. BMC medical research methodology 2010 Feb 10;10(1):13.

(29) Monroe S, Cai L. Evaluating Structural Equation Models for Categorical Outcomes: A New Test Statistic and a Practical Challenge of Interpretation. Multivariate behavioral research 2015 Nov 02;50(6):569-583.

(30) Clark DA, Bowles RP. Model Fit and Item Factor Analysis: Overfactoring, Underfactoring, and a Program to Guide Interpretation. Multivariate behavioral research 2018 Jul 04;53(4):544-558.

(31) Shi D, Lee T, Maydeu-Olivares A. Understanding the Model Size Effect on SEM Fit Indices. Educational and psychological measurement; 79(2019):310-334.

(32) Walfish S, McAlister B, O'Donnell P, Lambert MJ. An Investigation of Self-Assessment Bias in Mental Health Providers. Psychological reports 2012 Apr;110(2):639-644.

(33) Adams SA, Matthews CE, Ebbeling CB, Moore CG, Cunningham JE, Fulton J, et al. The effect of social desirability and social approval on self-reports of physical activity. American journal of epidemiology 2005;161(4):389-398.

(34) Portney LG, Watkins MP. Foundations of Clinical Research Applications to Practice. Third ed.: Pearson Education; 2014.

Table 1

Baseline Characteristics of Study Participants

Variable	*n*	*%*
Sex		
Male	61	11.8
Female	454	88.2
Profession		
ICU-nurse	65	12.5
Recovery nurse	4	0.80
ER-nurse	8	1.60
General nurse	383	74.5
Specialist nurse/nurse specialist	55	10.7
Education level[a]		
NLQF-level 4/4+	281	54.6
NLQF-level 6	227	44.0
NLQF-level 7	7	1.40

Note. *N* = 515. SD = standard deviation, ICU = intensive care unit, ER = emergency room.

[a]Adapted from "The Dutch Qualifications Framework (NLQF). Classification of Dutch qualifications in the NLQF and EQF. Available from: https://www.nlqf.nl/nlqf-niveaus. [accessed 5th November 2020]."

Table 2

Results of CFA by Model and Indices

Model	*RMSEA*	*CFI*	*TLI*
1) 1-factor model	0.193	0.924	0.911
2) 2-factor model	0.163	0.946	0.936
3) 1-factor model based on 2 factors with high factor correlation (>0.70)	0.163	0.946	0.936

Note. N = 515. CFA = Confirmatory Factor Analysis, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative fit index, TLI = Tucker Lewis Index.

Table 3

Factor Correlation Matrix

Factor	1	2
1	1.00	0.714
2	0.714	1.00

Note. N = 515. The extraction method was principal axis factoring with an oblique (Direct oblimin with Kaiser Normalization) rotation.

APPENDIX 1 – Dutch version of the NCRS

-- Volgende pagina --

74 **In welke mate zijn onderstaande stellingen op u van toepassing?**

	Volledig mee oneens	Mee oneens	Neutraal	Mee eens	Volledig mee eens
Ik weet hoe ik snel gegevens kan verzamelen over de gezondheid van een opgenomen patiënt	○	○	○	○	○
Ik beoordeel de verzamelde gegevens over de gezondheid van de patiënt adequaat	○	○	○	○	○
Ik stel afwijkingen vast in de verzamelde patiëntgegevens	○	○	○	○	○
Ik stel gezondheidsproblemen van een patiënt vast met de verzamelde afwijkende gegevens	○	○	○	○	○
Ik herken vroegtijdige tekenen of symptomen die wijzen op verslechtering van de gezondheidstoestand van een patiënt	○	○	○	○	○
Ik verklaar vroegtijdige tekenen of symptomen die wijzen op verslechtering van de gezondheidstoestand van een patiënt aan de hand van onderliggende mechanismen	○	○	○	○	○
Ik prioriteer patiënt problemen accuraat en maak daardoor problemen beheersbaar	○	○	○	○	○
Ik leg de mechanismen die ten grondslag liggen aan patiëntproblemen correct uit	○	○	○	○	○

-- Volgende pagina --

75 **In welke mate zijn onderstaande stellingen op u van toepassing?**

	Volledig mee oneens	Mee oneens	Neutraal	Mee eens	Volledig mee eens
Ik stel in samenspraak met de patiënt de juiste verpleegdoelen passend bij zijn of haar zorgproblemen	○	○	○	○	○
Ik voer passende verpleegkundige interventies uit voor de geconstateerde patiëntproblemen	○	○	○	○	○
Ik ben op de hoogte van alle beschikbare verpleegkundige interventies	○	○	○	○	○
Ik communiceer cruciale informatie over de conditie van de patiënt helder met artsen	○	○	○	○	○
Ik weet door de verstrekte informatie over de patiënt wat het beleid van de dokter wordt	○	○	○	○	○
Ik beoordeel of constateer nauwkeurig of de gezondheidstoestand van een patiënt is verbeterd	○	○	○	○	○
Ik weet welke vervolgstappen nodig zijn als de gezondheidstoestand van de patiënt niet verbetert	○	○	○	○	○

Als *kenmerk Leiding :* != L
Spring naar: vraag 83

-- Volgende pagina --