International Journal of Research in Medical Sciences
Abdillah IL et al. Int J Res Med Sci. 2017 Jan;5(1):283-289
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20164564

Original Research Article

Inter-rater reliability of wound care skills checklist in objective structured clinical examination

Ina Laela Abdillah[1], Intansari Nurjannah[2]*

[1]Undergraduate Student of School of Nursing, Faculty of Medicine, Universitas Gadjah Mada, Indonesia
[2]Associate Professor, Department of Basic and Emergency Nursing, Faculty of Medicine, Universitas Gadjah Mada, Indonesia

Received: 06 November 2016
Accepted: 03 December 2016

***Correspondence:**
Dr. Intansari Nurjannah,
E-mail: intansarin@ugm.ac.id

ABSTRACT

Background: The wound care skills checklist in objective structured clinical examination (OSCE) should be valid and reliable. Thus, the reliability test of the wound care skills checklist is needed. Purpose of the study was to identify the reliability of the wound care skills checklist.
Methods: This study is a descriptive non-experimental quantitative research with a cross-sectional study design. This study was conducted in the School of Nursing, Universitas Gadjah Mada, Indonesia. The number of respondents was 94 second-year students of this school of nursing. Inter-rater reliability was performed by 2 raters during OSCE. Kappa and Percent agreement (PA) were used to analyze the reliability of the checklist.
Results: Inter-rater reliability of the wound care skills checklist is categorized as good based on kappa value (0.7613) and acceptable based on PA value (89.36%). The results of the twenty-two item checklist were divided into five categories. Sixteen of the twenty-two items on the wound care skills checklist are included in the first category in which kappa category (≥0.41) and PA (>70%) are acceptable. One item is in the second category which has unacceptable value of kappa and PA, one item is in the third category which has low kappa value (0.3974) and high PA (89.36%), one item is in the fourth category which has a kappa value of 0, and three items are in the fifth category which has negative kappa value.
Conclusions: Inter-rater reliability of the wound care skills checklist OSCE in this nursing school can be categorized as good and acceptable.

Keywords: Checklist, Evaluation, Inter-rater reliability, Kappa, OSCE, Percent agreement

INTRODUCTION

A wound care skill is one of the competencies that have to be achieved by nursing students. The skills are obtained through skills laboratory learning and evaluated using a checklist in the Objective Structured Clinical Examination (OSCE) taking place in medical educational institutions.[1]

One of the educational institutions that applies the learning of wound care skills in the skills laboratory is the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada. As an instrument, the OSCE checklist should be valid and reliable. Reliability refers to the consistence variance of a measure.[2]

One of the methods to measure the reliability of an instrument is inter-rater reliability.[3] The number of OSCE-related researches are increasing, but are not in proportion to researches on the reliability of the OSCE instrument.[4] Based on the literature review conducted by the present researcher, there are no research publications

on the inter-rater reliability of the wound care skills checklist in the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada.

METHODS

Design

This research was a non-experimental quantitative descriptive research with a cross-sectional design.

Settings and participants

This research was conducted in June, 2015. It took place at the skills laboratory room of the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada. The population sample in this research were all second year students of the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada who took the OSCE of wound care skills at the end of the 4th semester, numbering 95 students.

The samples in this research were taken using total sampling. Inclusion criteria were all second-year students of the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada who took the OSCE on wound care skills. Exclusion criteria were students who were absent in the OSCE. One respondent was not willing to participate in this study, so that the total respondents were 94 students. Most respondents were female numbering 83 and the rest were male respondents numbering 11 people. The average age of the respondents in this study was 20 years old.

Ethical approval

A letter of application for ethical clearance was obtained from the Ethics Committee of the Faculty of Medicine, of Universitas Gadjah Mada-DR. Sardjito Hospital with the ethical clearance number of KE/FK/7/7/EC/2015. Every respondent who participated in this research was given explanation about the objectives and benefits of this research and signed an informed consent form agreeing to participate in the research. Confidentiality of respondents' information was guaranteed and all respondents who participated in this research were treated fairly.

Instrument

The instrument in this study is a wound care checklist. This checklist consists of 22 items divided by four phases and has three levels of measurement. The number of items at each stage in this checklist is 3 items in the pre-interactions phase, 3 items in the orientation phase, 11 items in the working stages, 4 items in the termination phase, and 1 item in the documentation stage.

The three level measurements namely are: performed perfectly (score of 2), performed but not perfectly (score of 1), and not performed at all (score of 0).

Data collection

Data were collected directly by two raters. The first raters were lecturers of the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada, who had been prepared by the skills laboratory coordinator to assess the wound care skills in the OSCE. The second rater was third-year students of the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada in accordance with the criteria in this research. The first rater consisted of 2 lecturers and the second rater consisted of 5 students. Raters were recruited before the OSCE took place and the raters were given explanation sheets and informed consent forms.

Prior to the OSCE the second raters were brought together with the first raters in order to avoid misperception in the assessment using the wound care skills checklist. Inter-rater's common perception was achieved through discussion. In addition, a second rater also simulated the assessment of wound care skills checklist before the OSCE took place. The OSCE of wound care skills took place for three days on June 15, 16 and 17, 2015. The first day was for groups 1, 2, and 3. The second day was for groups 4, 5, 6 and the third day for groups 7 and 8.

The first rater and the second rater assessed simultaneously when the second-year students performed the wound care skills in the OSCE. Respondents entered the laboratory skills room that had been determined as the place for the OSCE of wound care skills and in which there had been two assessors as the raters, namely the first rater and the second rater who assessed the performance of the students and there was a mannequin as the patient of the simulation. Students performed wound care skills for a maximum of 7 minutes and in accordance with the given case. There were four cases given in the OSCE of wound care skills. Students were asked to choose one of the cases and apply it in the wound care skills. The cases were made by the instructor of wound care skills in the same School of Nursing. Such cases would encourage cognitive, affective, and psychomotor skills of students in performing wound care skills, so that their competence in performing such skills could be assessed.

Data analysis

The assessment results of the first rater and second rater were analyzed using kappa test and percent agreement using STATA 9. The kappa value was interpreted using a scale interpretation by Altman's Benchmark Scale for the kappa.[5] Acceptable limit for kappa value in this research was ≥0.41, i.e. in the category of moderate, good, and very good based on the interpretation of kappa according to Altman's.[5] This is consistent with Cohen's suggestion in McHugh that the kappa value of 0.41 or higher is acceptable kappa value.[6] Acceptable limit of percent agreement is 70% or higher.[7]

RESULTS

Inter-rater reliability checklist of wound care skills

Checklist of wound care skills in this School

The assessment results for the inter-rater reliability of wound care skills checklist can be seen in the overall results of the assessment and the results of the assessment on each item. Table 1 shows the overall inter-rater's assessment for the wound care skills checklist after being categorized into pass (≥75) and fail (<75) in accordance with the assessment standards applied by the skills laboratory in this School.

The overall inter-rater's assessment for wound care skills checklist after being categorized into pass and fail indicated that there were 84 agreements and 10 disagreements. The wound care skills checklist on the whole had a kappa value with the category of good, 0.7613, ranging 0.60 to 0.80 based on the levels of reliability of Benchmark Scale according to Altman's.[5] The obtained value of percent agreement was 89.36%, categorized as acceptable (>70%).[7]

Table 1: Overall inter-rater's assessment.

	Rater 2				Kappa	Percent Agreement
Rater 1		**Pass**	**Fail**	**Total**	0.7613	89.36%
	Pass	58	1	59		
	Fail	9	26	35		
	Total	67	27	94		

Inter-rater reliability of wound care skills checklist items

The results of the calculation and interpretation of the kappa and percent agreement values on every item of wound care skills checklist are presented in Table 2. According to Table 2, the results of each item of wound care skills checklist range from poor to very good based on the interpretation of kappa according to Altman's.[5] Four items are categorized as poor, 2 items fair, 6 items moderate, 8 items good, and 2 items very good. Very good kappa value was at the documentation phase, namely the item of recording action performed and the patient's response and the item of feeling evaluation on the termination phase. Almost all the items on the checklist wound care skills have acceptable values of percent agreement, except the item of activity conclusion has an unacceptable value of percent agreement (<70%).

DISCUSSION

Inter-rater reliability of wound care skills checklist

The result showed a kappa value categorized as good and acceptable value of PA for total score of this instrument. It shows that overall the checklist has good reliability, so that the checklist is reliable or trustworthy to be used as an instrument to measure the wound care skills in the OSCE. One of the factors that can affect inter-rater reliability is the sharp difference in the data. According to the research of McHugh, if a variable has only two possible results and the results have sharp differences in the data, its reliability will most likely be high.[6] This is consistent with the results of the present research that the reliability of pass and fail variables is good due to differences in the data between rater 1 and rater 2. The calculation shows that of the total 94 respondents, there are 84 agreements and 10 disagreements between rater 1 and rater 2.

The results of this study are also consistent with the conclusion of Rushforth that the inter-rater reliability research often generates moderate or good value of inter-rater reliability.[8]

Inter-rater reliability of wound care skills checklist items

Overall wound care skills checklist has good reliability, but the results obtained on each checklist item indicate the value of kappa and percent agreement vary. Several items indicate different results between using kappa and percent agreement in determining the value of inter-rater reliability.

Based on Table 2, the results of each wound care skills checklist item is grouped into several groups based on acceptability of the values of kappa and percent agreement obtained. The data are grouped into 5 groups, namely acceptable kappa and percent agreement, unacceptable kappa and percent agreement, unacceptable kappa and acceptable percent agreement, kappa with 0 values, kappa with negative value.

Acceptable kappa and percent agreement

The items in this group have the kappa value of >0.40 and percent agreement value of >70. Based on Table 2, sixteen items can be categorised in this category. These items have a good value of kappa and percent agreement

can be accepted because items in this category have the characteristic of a good instrument which are specific, and has only one concept.[9]

Unacceptable kappa and percent agreement

This group has a kappa value of <0.41, or in the interpretation of kappa according to Altman's is in the category of poor and fair and the percent agreement value is <70%.

There is only one item in this group, namely the item of activity conclusion in the termination stage. The item indicates unacceptable values of kappa and percent agreement so it can be said that this item has a low reliability and needs improvement. This item only has one instruction for mentioning conclusion.

A good instrument needs to have one instruction at a time.[9] Although this item already followed a rule for good instrument in terms of instruction, however, this item may have more than one concept. How to conclude is something which every individual can have a different perception.

Thus this item may need to be revised by making it clear about what aspects need to be concluded. This revision is needed because low level of reliability is unacceptable in health research or clinical research.[6]

Table 2: Inter-rater reliability of wound care skills checklist items.

Items	Percent agreement	Kappa	Interpretation	
			Kappa[5]	Percent agreement[7]
Pre-Interaction Phase				
Verify Order/action	98.94%	0.0000	Poor	Acceptable
Prepare instruments	91.49%	0.7239	Good	Acceptable
Wash hands	80.85%	0.6767	Good	Acceptable
Orientation phase				
Greet and mention name	96.81%	-0.0144	Poor	Acceptable
Explain action (contract)	96.81%	-0.0144	Poor	Acceptable
Privacy	92.55%	0.7916	Good	Acceptable
Work phase				
Adjust the position of the client, give sheets	82.98%	0.5990	Moderate	Acceptable
Remove plaster and bandage using gloves/ tweezers and alcohol cotton ball	77.66%	-0.0930	Poor	Acceptable
Examine wound condition	76.60%	0.5098	Moderate	Acceptable
Wash hands	70.21%	0.5308	Moderate	Acceptable
Open sterile tools and prevent from contamination, pour antiseptic solution, add necessary tools and materials	86.17%	0.6207	Good	Acceptable
Use sterile gloves	88.30%	0.6165	Good	Acceptable
Clean the wound according to wound conditions and use the concept of moist wound dressing by maintaining sterility	80.85%	0.6871	Good	Acceptable
Cover the wound with sterile gauze in accordance with the wound condition	79.79%	0.6052	Moderate	Acceptable
Remove gloves	89.36%	0.3974	Fair	Acceptable
Fix the gauze with plaster	75.53%	0.4745	Moderate	Acceptable
Return the client to original position	75.53%	0.4860	Moderate	Acceptable
Termination phase				
Evaluation of feeling	93.62%	0.8437	Very Good	Acceptable
Activity conclusion	55.32%	0.3459	Fair	Unacceptable
Future contract	91.49%	0.8062	Good	Acceptable
Clean up tools and wash hands	74.47%	0.6132	Good	Acceptable
Documentation phase				
Record the action performed and patient's response	91.49%	0.8244	Very Good	Acceptable

Unacceptable kappa and acceptable percent agreement

There is one item which has a kappa value of <0.41 and a percent agreement value of >70%, namely the item of open gloves in the work phase. The item has a relatively high value of percent agreement and relatively low value of kappa (high agreement but low kappa). Some research suggests that items with the category of high agreement but low kappa can be caused by the nature of kappa, namely paradox that occurs due to the effect of prevalence.[10-12] The effect of the prevalence on the kappa paradox can be explained through the prevalence index obtained through the calculation of the differences of understanding divided by total subjects (n). The formula of prevalence index for a 2x2 contingency table is $\frac{|a-d|}{n}$.[13,14]

Another factor that affects kappa value is bias. The formula of bias index is for a 2x2 contingency table is $\frac{|b-c|}{n}$.[13,14] Bias index and prevalence index can be regarded as a paradox that affects the kappa value.[13,14]

A research conducted by Ludbrook states that a modification of the formula of bias index for 3x3 tables is $= \frac{|\sum UR - \sum LL|}{N}$.[15]

The formula of bias index for 3x3 tables in the item 15 namely open gloves is 0. The values of the bias index range from 0 to 1.[15]

Based on this statistic, the calculation of the value of bias index in the item of open glove is equal to the value of bias index in the item of evaluation of feeling, which is 0. However, the item of evaluation of feeling has a kappa value of 0.8437 and a percent agreement value of 93.62%, which means that kappa and percent agreement are categorized as acceptable, while the item of open glove has a kappa value of 0.3874 and a percent agreement value of 89.36%. This indicates that the two items, which have the same bias index and acceptable percent agreement, can produce different kappa values (ranging between acceptable and unacceptable).

This pattern is consistent with the view of Sim and Wright that low kappa value in kappa paradox is in accordance with low bias index.[13] Therefore, low kappa value (<0.41) and high percent agreement (>70%) in item 15 occur due to the kappa paradox.

Kappa with a value of 0

The item of verify order/action shows that the obtained kappa value is 0. A value of 0 indicates that Cohen's kappa is undefined or uninterpretable.[14] Kappa can have a value of 0 when the value of observed agreement is equal to the value expected agreement.[11,16,17] This result occurred because one rater constantly gave the same score for all respondents. 1 Rater 2 always gave a score of 2 in the item 1, so that the obtained value of Pa and Pe were similar, namely 98.94%.

Research on the reliability indicated that inter-rater reliability can generate a value of 0.18 The obtained inter-rater reliability value of 0 indicates that the rater is already accustomed to using the research instrument, i.e., a checklist.[1] The rater is also already familiar with the wound care skills checklist so that in giving a value in the item of verify order/action, rater 1 and rater 2 assumed that the respondents in performing the wound care skills would certainly verify order/action.

Kappa with negative value

The item of greeting and mention name and the item of explain the action (contract) at the orientation phase, as well as the item of remove the plaster and bandage using gloves/tweezers and alcohol cotton ball at the work phase have a negative kappa value. A negative kappa value indicates that the observed agreement or percent agreement is less than the expected agreement.[11,13,16,17] These results are consistent with the present research that the values of the observed agreement in the three items are 96.81%, 96.81%, and 77.66%, less than the expected agreement in the three items, namely 96.85%, 96.85%, and 79.56%.

The value of expected agreement plays an important role in calculating the kappa value. The values of expected agreement in the items of greet and mention name and the item of explain action (contract) at the orientation phase are very high and the item of remove plaster and bandage using gloves/tweezers and alcohol cotton ball are also quite high. The higher the expected value of the agreement over the limit values observed agreement, the lower the kappa value.[14]

A negative value may indicate poor agreement or disagreement between rater 1 and rater 2. The coefficient of negative value obtained in a research is usually small.[13] This is consistent with the results in the present research that the obtained negative values are -0.0144 and -0.0930. Low negative values, ranging less than -0.10, in general can be interpreted that there is no inter-rater agreement. A high negative kappa value indicates considerable disagreement between raters.[6]

Negative kappa value and high percent agreement value can be called the kappa paradox.[19] Negative values are categorized as poor.[19] If the kappa paradox occurs, it is necessary to consider the value of percent agreement. However, percent agreement also has a drawback because it is not able to calculate the value of expected agreement, while kappa is able to calculate the expected agreement.[6,20] In addition, according to Sim and Wright kappa is able to calculate "true agreement" and according to Joyce an assessment using kappa is able to provide more information than that using percent agreement.[13,21] Therefore, in this research the items included in the

category of kappa paradox indicate poor reliability and are suggested for revisions.

In general, the inter-rater reliability of the wound care skills checklist in the School of Nursing, Faculty of Medicine, Universitas Gadjah Mada is good, with only a few items indicating the unacceptable value of reliability and kappa paradox. Those items indicate that there is more than one instruction in a checklist item, for example in the item of greet and mention name and the item of remove plaster and bandage using gloves/tweezers and alcohol cotton ball. This is not in accordance with the characteristic of a good checklist item that is specific. Every checklist item should only describe a single concept, so that there should not be more than one instruction for each checklist item.[9] If two instructions exist in a checklist item, it should be split into two items.

Some other items that indicate unacceptable values of reliability or kappa paradox are the item of explain the action/contract, the item of open gloves, and the item of activity conclusion. The three items show that each item is less detailed in describing the instructions to be conducted. According to the view of Cazzell and Howe. However, the reliability of an item may be low because the item has less detail in describing instructions to be conducted, so that the item needs to be clarified in order to be more easily understood by students and raters.[22] Items in the checklist which fall in this category need to be divided into two items that have simple clear instruction.

Limitations

There are four cases used in this OSCE activity, which may influence how respondents respond to an activity in the wound care skills examination. The sample size also needs to be increased in the future research.

CONCLUSION

In conclusion, the inter-rater reliability value of the wound care skills checklist in Nursing Science Program Faculty of Medicine, Universitas Gadjah Mada is good and acceptable. Some items however need to be improved through revision and further tested in future research studies.

ACKNOWLEDGEMENTS

Authors would like to thank all respondents who made this research possible. Authors also would like to thank for School of Nursing Faculty of Medicine Universitas Gadjah Mada for financial support.

Funding: School of Nursing Faculty of Medicine Universitas Gadjah Mada
Conflict of interest: None declared
Ethical approval: The study was approved by the Institution Ethic Committee Faculty of Medicine Universitas Gadjah Mada

REFERENCES

1. Houghton C, Casey D, Shaw D, Murphy K. Staff and students' perceptions and experiences of teaching and assessment in Clinical Skills Laboratories: Interview findings from a multiple case study. Nurse Education Today. 2012;32(6):e29-34.
2. Schwab DP. Research methods for organizational studies: Psychology Press. 2013.
3. Peyré S, Peyré C, Hagen J, Sullivan M. Reliability of a procedural checklist as a high-stakes measurement of advanced technical skill. Amer J Surg. 2010;199(1):110-4.
4. Patrício M, Julião M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. Medical teacher. 2009;31(2):112-24.
5. McCray G, editor. Assessing inter-rater agreement for nominal judgement variables. Language Testing Forum. 2013.
6. McHugh M. Interrater reliability: the kappa statistic. Biochemia Medica. 2012;22(3):276-82.
7. Osborne JW. Best practices in quantitative methods: Sage. 2008.
8. Rushforth H. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. Nurse Education Today. 2007;27(5):481-90.
9. Medical Council of Canada. Guidelines of the development of objective structured clinical examination (OSCE) cases. 2013; Available from: http://mcc.ca/ wp-content/ uploads/ osce-booklet-2014.pdf.
10. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. Journal of clinical epidemiology. 1990;43(6):543-9.
11. Viera A, Garrett J. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360-3.
12. Cargo M, Stankov I, Thomas J, Saini M, Rogers P, Mayo-Wilson E, et al. Development, inter-rater reliability and feasibility of a checklist to assess implementation (Ch-IMP) in systematic reviews: the case of provider-based prevention and treatment programs targeting children and youth. BMC medical research methodology. 2015;15(1):1.
13. Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical therapy. 2005;85(3):257-68.
14. Xie Q. Agree or Disagree? A Demonstration of An Alternative Statistic to Cohen's Kappa for Measuring the Extent and Reliability of Agreement between Observers. 2013.

15. Ludbrook J. Detecting systematic bias between two raters. Clin experi pharma physio. 2004;31(1-2):113-5.
16. Gisev N, Bell J, Chen T. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm. 2013;9(3):330-8.
17. Kvålseth T. Measurement of Interobserver Disagreement: Correction of Cohen's Kappa for Negative Values. J Proba Statis. 2015;2015.
18. Krippendorff K. Agreement and information in the reliability of coding. Comm Met Meas. 2011;5(2):93-112.
19. Cunningham M. More than just the kappa coefficient: A program to fully characterize inter-rater reliability between two raters, SAS Global Forum: Statistics and data analysis. Retrieved Sept. 2009;21:2010.
20. Graham M, Milanowski A, Miller J. Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. Online Submission. 2012.
21. Joyce M, editor. Picking the best intercoder reliability statistic for your digital activism content analysis. Digital Activism Research Project: Investigating the Global Impact of Comment Forum Speech as a Mirror of Mainstream Discourse; 2013.
22. Cazzell M, Howe C. Using objective structured clinical evaluation for simulation evaluation: Checklist considerations for interrater reliability. Clinical simulation in Nursing. 2012;8(6):e219-e25.

Cite this article as: Abdillah IL, Nurjannah I. Inter-rater reliability of wound care skills checklist in objective structured clinical examination. Int J Res Med Sci 2017;5:283-9.