VALIDITY AND RELIABILITY OF NURSING INTERVENTION CLASSIFICATION: SELF-CARE ASSISTANCE ON PATIENTS WITH STROKE

Intansari Nurjannah*

BNJ BELITUNG NURSING JOURNAL

ORIGINAL RESEARCH:
RESEARCH METHODOLOGY PAPER

VALIDITY AND RELIABILITY OF NURSING INTERVENTION CLASSIFICATION: SELF-CARE ASSISTANCE ON PATIENTS WITH STROKE

Intansari Nurjannah*

Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada

*Correspondence:
Intansari Nurjannah, S.Kp., MN. Sc., Ph.D
Department of Mental Health and Community Nursing, Faculty of Medicine, Public Health and Nursing Universitas Gadjah Mada
Gedung Ismangoen, Sekip
Bulaksumur Yogyakarta, Indonesia
Phone : +62 274 545674
Fax: +62 274 631204
Email: intansarin@ugm.ac.id

Abstract

Background: Nursing intervention is part of nursing process. The accurateness of intervention needs to be explored through an effort to measure validity and reliability of the intervention.
Objectives: This study aimed to investigate the validity and reliability of four Nursing Intervention Classifications (NICs) of Self-Care Assistance (SCA) on patients with stroke.
Methods: Validity measurement involved 4 experts, while reliability involved 7 samples for each NIC. Validity was analyzed using content validity index (I-CVI and S-CVI), while reliability was analyzed using kappa and percent agreement.
Results: Sixteen activities of NICs (I-CVI score less than 0.78) were eliminated and two activities considered not applicable. The results of reliability were above 0.85 kappa value with 85% of percent agreement.
Conclusion: Elimination of not valid activities increased reliability.

KEYWORDS
validity; reliability; nursing intervention classification

INTRODUCTION

Nursing intervention is part of the nursing process, which needs to be continuously developed. There are several standards of nursing interventions that can be used by nurses. The standard of nursing intervention will describe the nursing care activity and is useful for communication across settings (Butcher, Bulechek, Dochterman, & Wagner, 2018). One of the nursing intervention standards is Nursing Intervention Classification (NIC). This classification has also been reviewed and updated continuously and consists of label name, a definition, a set of activities and background reading for each label of intervention (Butcher et al., 2018). The essence of each intervention is in the label name and the definition of the intervention while the activities can be modified as long as it is congruent with the

definition of the intervention (Butcher et al., 2018). Although its activities are not the essence of an intervention, however research needs to be conducted to find out which activities fit to a specific population.

This type of research is called descriptive research in which the goals are to validate nursing interventions in a specific population, to find out the most common core interventions and it also can be used to measure nursing workload (Butcher et al., 2018). The more precise activities of NIC to be used for each patient population, the easier for nurses to select which activity has to be applied to the patient. In addition, research needs to be conducted to determine the list of "compulsory activities" and

"additional activities" which will help refine the classification. Before further research can be done, validity and reliability studies need to be conducted to ensure the integrity of a study's results (DeVon et al., 2007; Souza, Alexandre, & Guirardello, 2017).

NIC has been widely used and translated into different languages (Butcher et al., 2018) and because of that, there is a risk for reducing the essential meanings of NIC in a different language. In this case, translation of NIC into different languages needs to be checked for validity and reliability as well, to find out whether NIC is applicable or not to be used in a country different from an English-speaking country.

Content validity measurements have the highest priority to be done if we compare them to other validity measurements, because this type of validity describes which items adequately represent the content of a domain (Zamanzadeh et al., 2015). Content validity is related to the number of underlying dimensions, the number of items, and the theoretical framework (DeVon et al., 2007). The content validity study will help to reduce the revision of the measurement in the future with fewer resources for evaluation phase (Rubio, Berg-Weger, Tebb, Lee, & Rauch). The clarity of an item will also be addressed through content validity study, and it can also be used to assess other psychometric properties (Rubio et al.), and will help to make the item able to be more easily measured (Wynd, Schmidt, & Schaefer, 2003). Content validity can be measured by both a qualitative approach (reviewing by experts) and a quantitative approach (by using content validity index) (Souza et al., 2017). Using a panel of experts provides constructive feedback about the quality of the newly developed measure and objective criteria with which to evaluate each item (Rubio et al., 2003).

Measuring reliability of NICs items is also necessary to be done. Reliability is defined as measurement precision or the ability to reproduce the scores obtained from measurement (van der Vlauten, 2000). One type of measuring reliability is by using interrater reliability, which is a method to measure items where observation methods and two raters are involved (Gwet, 2014). Since there is no best method of analysis to measure interrater reliability, Graham et al. (2012) suggested to use two different methods of analysis which are kappa and percent agreement at the same time. Several researches have been conducted using these two methods of analyses (Morris, MacNeela, Scott, Treacy, Hyde, O'Brien, et al., 2008; Peyré, Peyré, Hagen, & Sullivan, 2010; Tidsrand & Horneij, 2009), which make the results of study stronger and more reliable.

METHODS

Study design
This research was a descriptive study to measure the validity and reliability of four NICs SCA: Bathing, Feeding, Dressing/Grooming, and Toileting on stroke patients. Content validity was selected for validity analysis and interrater reliability was selected to measure reliability. Data collection for content validity was conducted in November-December

2017 and in January 2018 for interrater reliability measurement. Data for reliability measurements were collected from the stroke unit at one hospital in Yogyakarta, Indonesia.

Validity
Polit and Beck (2006) mentioned that 3 experts minimum are needed in the assessment process for content validity of a study. In this study there were four experts (two experts from nursing academicians and two experts from nursing clinicians) were involved in the process of measuring content validity. Criteria for selecting these experts were their number of publications or their work experience (Rubio et al.). Ideally, an expert panel consists of professional experts and lay experts (Rubio et al.), however in this study, only panel experts with professional criteria were used for measuring content validity. Content validity can be measured qualitatively (expert judgment) and quantitatively (by using I-CVI and S-CVI) (Souza et al., 2017). This research used both measurements as Wynd et al. (2003) mentioned that qualitative measurement depends only on judgment, logic and reasoning of experts in the domain content. Content Validity Index for Item (I-CVI) refers to experts' review of the items (Polit & Beck, 2006). Experts used four choice scale which are 1 = not relevant, 2 = somehow relevant, 3 = quite relevant, 4 = high relevant toward the level of relevance, accuracy, clarity and ambiguity of item in each NICs. Scores for I-CVI were considered as relevant if the score was ≥ 0.78 when ≥ 3 experts were involved (Polit, Beck, & Owen, 2007). Items are also considered as valid if the overall scale (S) score of S-CVI was ≥ 0.80 (Polit & Beck, 2006). Based on I-CVI and S-CVI values, several items of activities in the instrument were eliminated before it was used for reliability measurement.

Reliability
The sample for interrater reliability was calculated using a formula 2 x 2 table, with confidence level 80%, and alpha 0.05, and the result show that the recommended number of samples for kappa test was 7 for each NICs (Bujang & Baharum, 2017). Raters were two student nurses who have passed their internship period. They were trained by the researcher before they collected the data. In this study, interrater reliability was selected to measure reliability. Kappa and percent agreement were used to analyze the data. Based on the Benchmark Scale for the Kappa (Altman, 1991) cited in McCray (2013), the score > 0.80 was considered very good, or almost perfect (Landis and Koch, 1977 cited in McCray, 2013), while Fleiss (1981) cited in McCray (2013) argued that value > 75 is excellent. For analysis using percent agreement, the result can be divided into two categories, which are accepted and unaccepted. Several literatures mention that the PA value $\geq$ 70% is accepted (Osborne, 2008); Stemler, 2004 in Morris et al., (2008). In the process of reliability data collection, the researcher limited the number of observations by raters to only two observations each day.

Instrument translation
Instrument for this data collection was originally in English. Researcher then checked the translation of Indonesian publication for this instrument as this instrument can be found

translation book in Bahasa Indonesia (Bulechek, Butcher, Dochterman, & Wagner, 2013).

Ethical consideration
Ethical approval was obtained from the Ethics Committee of the Faculty of Medicine, Universitas Gadjah Mada. The number of ethical approval was Ref: KE / FK / 1121 / EC / 2017. Respondents were asked to sign the informed consent when they were willing to participate in this study.

RESULTS

The results of study showed that there were as many as 16 activities from a total of 72 activities of NICs that have I-CVI scores lower than the cut-off (less than 0.78). There were two activities in SCA: Feeding which are: "Use a cup with a large handle, if necessary" and "Use unbreakable and weighted dishes and glass, as necessary" with score 1 in accuracy and clarity (see **Table 1**). However, as suggested by expert panels, these two activities were not applicable to be applied in a hospital setting. The researcher then eliminated these activities as well as 16 other activities, which had low I-CVI scores from the list activities that will be used for reliability measurement.

Table 1 Items of activities from four NICs SAC Bathing, Feeding, Dressing/Grooming, Bathing that were eliminated

NIC	Activities needing to be eliminated	I-CVI score			
		Relevance	Accuracy	Clarity	Ambiguity
Self-Care Deficit Bathing	Facilitate maintenance of patient's usual bedtime routines, pro sleep cues/props, and familiar objects (e.g., for children, favorite blanket/toy, rocking, pacifier, or story; for adults, book to read or a pillow from home), as appropriate	.5	.5	.5	.5
	Encourage parent/family participation in usual bedtime ritual, as appropriate	.25	.25	.25	.25
Self-Care Assistance: Feeding	Create a pleasant environment during meal time (e.g., put bedpans, urinal, and suctioning equipment out of sight)	.5	.75	.5	.5
	Encourage patient to eat in dining room, if available	.25	.5	.5	.25
	Provide social interaction, as appropriate	.5	.5	.5	.25
	Provide adaptive devices to facilitate patient's feeding self (e.g., long handles, handle with large circumference, or small strap on utensil), as needed	.5	.75	.75	.5
	Use a cup with a large handle, if necessary	.75	1	1	.75
	Use unbreakable and weighted dishes and glass, as necessary	.75	1	1	.75
	Provide frequent cueing and close supervision, as appropriate	.25	.25	.25	.25
Self-Care Assistance: Dressing/Grooming	Use extension equipment for pulling on clothing, if appropriate	.5	.5	.25	.25
	Offer to launder clothing, as necessary	.25	.75	.75	.5
	Offer to hang up clothing or place in dresser	.25	.5	.5	.25
	Offer to rinse special garments, such as nylons	.	.25	.25	.25
	Provide fingernail polish, if requested	.	.5	.5	.25
	Provide makeup, if requested	.25	.5	.5	.25
	Facilitate assistance of a barber or beautician, as necessary	.5	.75	.75	.5
Self-Care Assistance: Toileting	Consider patient's response to lack of privacy	.5	.75	.75	.75
	Institute a toileting schedule, as appropriate	.75	.75	.75	.75

The results of this study showed that the S-CVI/Ave Approach score of relevance, accuracy, clarity and ambiguity increased after several items, which have low score of I-CVI, were eliminated from the list of activities in each NIC. The kappa and percent agreement values were consistent and scored very high as can be seen in **Table 2**.

Table 2 Content validity score and interrater reliability score of NICs SCA: Bathing, Feeding, Dressing/Grooming and Toileting

NICs	Activities			S-CVI/Ave Approach score								Kappa	PA
				Relevance		Accuracy		Clarity		Ambiguity			
	Total	Eliminated	*)	**)	*)	**)	*)	**)	*)	**)		
Self-Care Assistance: Bathing	14	2	.83	.9	.82	.89	.80	.87	.82	.89	.89	91.67%
Self-Care Assistance: Feeding	25	7	.55	.98	.55	.97	.64	1	.47	.97	.95	98%
Self-Care Assistance: Dressing/ Grooming	19	7	.69	.95	.78	.93	.77	.93	.71	.93	.87	94%
Self-Care Assistance: Toileting	14	2	.91	.95	.91	.93	.92	.95	.89	.91	.93	85.71%

*) Score S-CVI before several activities in NICs were eliminated.
**) Score S-CVI after several activities in NICs were eliminated.

DISCUSSION

Results show that as many as 18 activities from four NICs were eliminated. In NIC SCA: Bathing, two activities which were eliminated can be considered as not fitting with the definition of NIC Self-Care Deficit: Bathing as those activities are related to sleep activities. Beside this reason, those items were also difficult to be applied because in this setting the hospital system does not allow a family to accompany the patient at all times, and the family has to stay outside of the patient's room. Family can only accompany the patient during the visitor time period.

In NIC SCA: Feeding, there were 7 activities that needed to be eliminated consisting of 5 activities with low score of I-CVI and two activities with adequate score in I-CVI but experts considered these two activities were not applicable to be applied in the hospital setting. Those two activities were "Use a cup with a large handle, if necessary" and "Use unbreakable and weighted dishes and glass, as necessary".

Most of the other activities related to items provided by a facility such as specific device – which may not available in the hospital, such as a specific room for eating (dining room), which is uncommon to be provided for patient in this hospital. The reason why an activity such as to "provide frequent cueing and close supervision" was likely not to be applied is because nurses may have high workload. Another item eliminated was "Create a pleasant activity during meal time" which also has low value of I-CVI because in this hospital, providing meals and setting the meal is non-nursing tasks. The activities to provide social interaction during mealtime also needed to be eliminated because most stroke patients are unable to enjoy their social interaction activity because of their condition. This is supported by a study ([Olshansky et al., 2012](#)) in which content validity was influenced by development and appropriate content for diverse individuals.

In NIC SCA: Dressing/Grooming, there were also 7 activities needing to be eliminated. Again, the main reason was mostly because there is no facility in this hospital to apply this activity. Several other activities were uncommon to be conducted, which are: provide fingernail polish and makeup, even though this also may be influenced by patients' condition, as they may need more priority activities than these activities. It also supported by a study which content validity was influenced by content words or terms that do not reflect common usage in the population being studied and also the depth and variation of the meaning of a specific term ([Olshansky et al., 2012](#)).

In NIC SCA: Toileting, the two activities have low content validity and needed to be eliminated because there was no clear method to assess patient's response to lack of privacy considering that stroke patients mostly have difficulty to communicate. "Institute a toileting schedule activity" also was eliminated because of patients' condition for example if they use catheter for their elimination, and again it is related to the populations being studied ([Olshansky et al., 2012](#)).

An important note needs to be made regarding the type of activities that needed to be eliminated. From 18 activities that needed to be eliminated, there are as many as 14 activities containing a modifier such 'as appropriate' (5 item), 'as necessary' (3 item), 'if requested' (2 item), and one each for 'as needed', 'if appropriate', 'if available' and 'if necessary'. There is no clear direction of the function of this modifier. Those modifiers may need to be categorized into 'compulsory/core activity' or 'additional activity', and then the usage of NIC can be more specific and easier to be used by nurses. Although elimination of 18 activities from four NICs seems to be many, however, it may help to refine the NICs from abstract concepts to measurable indicators ([Wynd et al., 2003](#)).

The second part of this discussion is related to the reliability of four NICs. Results show that the reliability of these four NICs

was consistent using both kappa value and percent agreement. Even though the reliability of these NICs is very good and accepted, however, we need to note that the results of reliability measurement were high when several items (total 18 items) in the NICs were eliminated. It is supported by Graham et al. (Graham et al., 2012), and Neuendorf (2002) that stated an instrument contributes to influence the score of interrater reliability. The easier instrument can be understood by a rater then the rater can be more consistent to provide a judgment toward each item (Graham et al., 2012). Roach (2006) also mentioned that the type of instrument would also influence the reliability estimates. In this study, however, the requirement to eliminate several items in NIC was related to the specific population, which were stroke patients. Those items may still be kept in the NIC list of activities if the NICs were targeted to a different patient population and setting, for example in another community setting.

Other factors that influence reliability are: training for rater, selection of rater, and whether raters were exhausted or not (Graham et al., 2012; Neundorf, 2002). Those criteria of factors that contribute to the value of interrater reliability also have been fulfilled in this research, such as training for raters. Graham et al. (2012) stated that training for raters is important to increase the value of interrater reliability. This training is necessary to reduce the number of variations in interpreting data in the instrument (McHugh, 2012). The period of training also is important, as rater training is better if it is around 1-2 hours (Graham, 2012) and can be divided into several sessions, as necessary (Neundorf, 2002).

Another factor that influences the interrater reliability of NICs is selection of raters, which is where the researcher ensures that raters have the same experience regarding the topic (Roach, 2006). In this study, raters have the same previous experience as they were nursing students who were in the same level and have passed their internship at the same time. Several researches show a positive relationship between rater expertise and the ability to judge an item accurately (Kozlowski & Kirsch, 1987) and Kozlowski & Kirsch, 1987; Smither, Barry, & Reilly, 1989 cited in Graham et al., (2012). However, there is still debate regarding who should be the rater, the qualification of rater and what rater experience, which will influence the value of interrater reliability (Rushforth, 2007). Another factor that influences the value of interrater reliability was rater's fatigue. Raters' schedule to do observation needs to be considered to be rational and realistic, and does not overload their workload (Neundorf, 2002). In this study, the researcher anticipated rater's fatigue by limiting observation activity into two observations only per day.

CONCLUSION

Although several activities in NICs SCA: Bathing, Feeding, Dressing/Grooming, Toileting, needed to be eliminated, this research was conducted in a specific population, in which several other activities may still be kept as activities in each NICs. Several activities in the NICs may refer to a specific population and culture in which there may be a need to divide activities in NICs into 'compulsory/core activities' and 'additional activities'. Those NICs that have a good value of content validity lead to very good result in their reliability scores.

ORCID
Intansari Nurjannah http://orcid.org/0000-0002-5406-8727

REFERENCES

Bujang, M., A, & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, Biostatistics and Public Health, 14*(2).

Bulechek, G, M, Butcher, H., K, Dochterman, J., M, & Wagner, C., M. (2013). *Nursing Intervention Classification (NIC) Dalam Bahasa Indonesia*. Missouri: Elsevier Mosby.

Butcher, H. K., Bulechek, G. M., Dochterman, J. M. M., & Wagner, C. (2018). *Nursing Interventions classification (NIC)-E-Book*: Elsevier Health Sciences.

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., . . . Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship, 39*(2), 155-164.https://doi.org/10.1111/j.1547-5069.2007.00161.x

Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. *Online Submission*.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*: Advanced Analytics, LLC.

Kozlowski, S. W., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. *Journal of Applied Psychology, 72*(2), 252. https://doi.org/10.1037/0021-9010.72.2.252

McCray, G. (2013). *Assessing inter-rater agreement for nominal judgement variables*. Paper presented at the Language Testing Forum.

Morris, R., MacNeela, P., Scott, A., Treacy, P., Hyde, A., O'Brien, J., . . . Drennan, J. (2008). Ambiguities and conflicting results: the limitations of the kappa statistic in establishing the interrater reliability of the irish nursing minimum data set for mental health: a discussion paper. *International Journal of Nursing*

Studies, 45, 645-647. https://doi.org/10.1016/j.ijnurstu.2007.07.005

Neundorf, K. (2002). The content analysis guidebook. *Thousands Oaks: Sage*.

Olshansky, E., Lakes, K. D., Vaughan, J., Gravem, D., Rich, J. K., David, M., . . . Cooper, D. (2012). Enhancing the construct and content validity of rating scales for clinical research: Using qualitative methods to develop a rating scale to assess parental perceptions of their role in promoting infant exercise. *The International Journal of Educational and Psychological Assessment, 10*(1), 36.

Osborne, J., W. (2008). *Best practice in quantitative methods. California*: Sage.

Peyré, S., Peyré, C., Hagen, J., & Sullivan, M. (2010). Reliability of a procedural checklist as a high-stakes measurement of advanced technical skill. *The American Journal of Surgery, 199*(1), 110-114. https://doi.org/10.1016/j.amjsurg.2009.10.003

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467. https://doi.org/10.1002/nur.20199

Roach, K., E. (2006). Measurement of health outcomes: reliability, validity and responsiveness. . *Journal of*

Prosthetics and orthotics, 18(IS), 8-12. https://doi.org/10.1097/00008526-200601001-00003

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94-104. https://doi.org/10.1093/swr/27.2.94

Rushforth, H. (2007). Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse Education Today, 27*(5), 481-490. https://doi.org/10.1016/j.nedt.2006.08.009

Souza, A. C. d., Alexandre, N. M. C., & Guirardello, E. d. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde, 26*(3), 649-659.

Tidsrand, J., & Horneij, E. (2009). Inter-rater reliability of three standardized functional tests in patient with low back pain. *BMC Musculoskeletal Disorders, 10*(5), 1471-2472. https://doi.org/10.1186/1471-2474-10-58

van der Vlauten, C. (2000). Validity of final examinations in undergraduate medical training. *BMJ, 321*(7270), 1217. https://doi.org/10.1136/bmj.321.7270.1217

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*(5), 508-518. https://doi.org/10.1177/0193945903252998

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences, 4*(2), 165. https://doi.org/10.15171/jcs.2015.017