Editorial

BNJ

Belitung Nursing Journal
Volume 7(5), 356-360

https://doi.org/10.33546/bnj.1927

Establishing appropriate sample size for developing and validating a questionnaire in nursing research

Joko Gunawan[1,2*], Colleen Marzilli[3], and Yupin Aungsuroch[1*]

Abstract

The number thirty is often used as the sample size in multiple questionnaires and identified as appropriate for validation of nursing research. However, this is not the best tool or strategy for sample size selection for development and validation, and this often causes immediate rejections of manuscripts. This editorial aims to provide an overview of the appropriate sample size for questionnaire development and validation. The article is the amalgamation of technical literature and lessons learned from our experiences in developing, validating, or adapting a number of questionnaires.

Keywords

questionnaire; validation; instrument development; sample size; nursing research

The significance of this editorial is the rejection rate (>85%) of the research articles submitted to the Belitung Nursing Journal (BNJ). The most common reasons for rejection are related to the sample size for instrument development and validation. Therefore, it is important to provide an explanation of the rationale for the appropriate sample size so it is clearly established.

The majority of the research articles submitted to BNJ use questionnaires. A questionnaire refers to the main instrument for collecting data in survey research. Basically, it is a set of standardized questions, often called items, which follow a fixed scheme in order to collect individual data about one or more specific topics (Lavrakas, 2008). In addition, the questionnaire is either developed by the researchers or modified from existing instruments.

Although BNJ's guideline clearly states that the author(s) should clearly describe the details of the questionnaires used for data collection, whether they develop, adopt, adapt, modify, or translate the instrument, many authors are confused about the terms and find it difficult to calculate or decide the appropriate sample size. Often, authors used a sample size of 30 as a golden rule number for all validation scenarios. Therefore, this editorial aims to provide an overview of the appropriate sample size used to develop and validate a nursing research questionnaire. This editorial is not a systematic review, but rather it is a technical literature amalgamation of lessons learned from our experiences in questionnaire development, validation, and adaptation. For the sake of consistency, we use the term "questionnaire" instead of scale, instrument, or inventory. In this article, we describe sample size based on the stages of questionnaire development and adaptation.

Sample Size for Questionnaire Development

The questionnaire development refers to a process of developing reliable and valid measures of a construct in order to assess an attribute of interest. Typically, the instrument development has two phases (DeVellis, 1991): instrument construction and psychometric evaluation. Meanwhile, from the perspective of mixed-methods research designs, instrumentation consists of qualitative and quantitative strands. However, both perspectives are similar because in the instrument construction stage, an item pool is generated, which may involve expert interviews

[1]Faculty of Nursing, Chulalongkorn University, Bangkok, Thailand
[2]Belitung Raya Foundation, Manggar, East Belitung, Bangka Belitung, Indonesia
[3]The University of Texas at Tyler, School of Nursing, 3900 University Blvd., Tyler, TX 75799, USA

Corresponding authors:
Joko Gunawan, S.Kep. Ners, PhD & Yupin Aungsuroch, PhD, RN
Faculty of Nursing, Chulalongkorn University
Borommaratchachonnani Srisataphat Building, Rama 1 Rd, Pathumwan, Bangkok 10330, Thailand
Email: jokogunawan@belitungraya.org | yaungsuroch@gmail.com

Article Info:
Received: 5 October 2021
Revised: 19 October 2021
Accepted: 28 October 2021

E-ISSN: 2477-4073 | P-ISSN: 2528-181X

that are considered qualitative in nature. In comparison, psychometric testing is regarded as a quantitative stage consisting of a questionnaire survey with large samples. However, in this article, we do not discuss the philosophical underpinnings of the two perspectives, rather the editors describe the sample size needed in each stage of instrument development.

In the instrument construction phase, samples may be needed to generate an item pool in order to get input from experts. It is essential for a study to bring a specific context, culture, or a dearth of published articles for item generations. The number of samples for interviews varies, from one to 50, depending on the scope of the study, the nature of the topic (i.e., complexity, accessibility), the quality of data, and the study design (Morse, 2000). In addition, researchers can also utilize the Delphi technique with a series of rounds, typically three rounds, to reach a consensus among experts as they review, discuss, accept, or reject items. The number of samples for the Delphi technique also varies, from 10 to 100 or more (Akins et al., 2005). However, it is noteworthy that expert interviews or the Delphi technique are not a must in developing an item pool. The researchers can choose using literature review, expert interviews, or the Delphi technique alone, or researchers can use a combination of a literature review and interviews. There is no golden standard for this stage as long as an explicit rationale is provided.

The samples are also needed in step 4 (instrument validation) and step 5 (pretesting or piloting the instrument) for researchers to engage in the instrument construction phase (See **Figure 1**). Therefore, although the researchers do not conduct an interview for item generation, they still need to find experts for validating instruments, especially for measuring the Content Validity Index (CVI). The recommended number of experts to review a tool varies from two to 20 individuals (Armstrong et al., 2005). At least five people are suggested to check the instrument to have sufficient control over chance agreement (Zamanzadeh et al., 2015). It is important to note that in the pretesting, or the pilot testing of the questionnaire, 15-30 subjects are recommended (Burns & Grove, 2005). This pilot testing is necessary before further examination utilizing a bigger sample size or phase II evaluation, or the psychometric properties evaluation, to ensure the construct validity and reliability of the instrument. The instrument will not be considered valid without the psychometric properties stage, especially when developing a new questionnaire.

To ensure the psychometric properties, or validity and reliability, of the newly developed questionnaire, factor analysis is one common tool. Conducting an Exploratory Factor Analysis (EFA) only or both an EFA and a Confirmatory Factor Analysis (CFA) are two options for factor analysis, and either of the two options is acceptable and viable for questionnaire development. It is noted that EFA is used for instruments that have never been tested before (to explore items and factor structures). In contrast, CFA is used for tested instruments to confirm and validate the items and factor structures. In other words, EFA is used to illustrate or to determine underlying latent variables or factors, and CFA is to check whether it fits reality (Knekta et al., 2019). Given these two different tools, the EFA and CFA must be conducted on different datasets; otherwise, overfitting is likely. If we try to verify the factor(s) we discovered with EFA using the same data, CFA results will most likely give good fit indices because the same data will tend to conform to the structure(s) of the scale, which is discovered with EFA.

It is also noted that the factor analysis literature for both EFA and CFA contains a variety of recommendations regarding the minimum or appropriate sample size. Although both methods have different purposes and criteria, there is no golden standard to differentiate the sample size between the two methods. Additionally, most of the recommendations are often overlapping with each other, and in some cases, the recommendations may seemingly be contradictory. We provide a summary of the recommendations in **Table 1**, which can be grouped into the recommended sample size, the recommended item-to-response ratios, and the recommended estimated parameter-to-sample ratios.

The recommended sample size for factor analyses varies from 50 to more than 1000 samples, while the recommended item-to-response ratio is from 1:3 to 1:20. Also, the estimated parameter-to-sample ratio is from 1:5 to 1:20. The parameter-to-sample ratio is mostly used for a study with Structural Equation Modelling (SEM), of which CFA is a part. However, all suggestions are based on different perspectives. For EFA, the sample size is according to replicable factor structures, stable item/factor loadings, or strong data. Strong data includes high communalities, no cross-loadings, strong primary loadings per factor, the nature of the data, number of factors, or number of items per factor (Boateng et al., 2018; Kyriazos, 2018). While for CFA, or SEM in general, sample size depends on study design, such as cross-sectional vs. longitudinal; number of factors; number of relationships among indicators; the magnitude of the item-factor correlations; indicator reliability; the data scaling or categorical versus continuous; estimator type; parameters per measured variable number; the ratio of cases to free parameters; standard errors; missing data levels and patterns; and model complexity (Brown, 2015; Boateng et al., 2018; Kyriazos, 2018).

From **Table 1**, the reader may see that no single recommended sample size or item-to-response ratio fits all. However, a smaller sample size when all other things are equal is not as desirable as a large sample size because a larger sample lends itself to lower measurement errors, accuracy of population estimates, stable factor loadings, generalizability results, and model fit.

However, the sample size is always constrained by resources available, and more often than not, instrument development can be challenging to fund. Therefore, the minimum number of appropriate sample size in each research article should be evaluated individually. It is noteworthy that 30 subjects are not described in any factor

analysis literature for psychometric properties, except in pilot testing. Even 50 subjects are less likely to be recommended, as it will usually result in very unstable estimates, especially with psychological, social science, or nursing science data. However, if it is used in very accurate chemical measurements, 50 subjects may be appropriate. The researchers should provide clear rationale when they select the minimum criteria of the sample size. For example, if the questionnaire is specifically developed for patients with a specific disease, a bigger sample size is not applicable due to a limited number of patients.

Table 1 A variety of recommended sample sizes for factor analyses

A variety of recommended sample sizes for factor analyses	
Of sample size	
50	Barrett and Kline (1981)
100	Gorsuch (1983), Kline (1994)
≥150	Hutcheson and N. (1999)
≥150 - ≤180	Mundfrom et al. (2005)
200	Guilford (1954)
≥200	Hair et al. (2010)
250	Cattell (1978)
200 – 300	Guadagnoli and Velicer (1988), Comrey (1988)
300	Clark and Watson (2016)
400	Aleamoni (1976)
100 - >1000	Mundfrom et al. (2005)
50 = very poor, 100 = poor, 200 = fair, 300 = good, 500 = very good, ≥1,000 = excellent	Comrey and Lee (1992)
Of item to response ratio (*p*: *N*)	
1:3 to 1:6	Cattell (1978)
1:4	Rummel (1988)
1:5	Gorsuch (1983), Hatcher (1994)
1:10	Nunnally (1978), Everitt (1975), Watson and Thompson (2006)
1:3 to 1:20	Mundfrom et al. (2005)
Of estimated parameter to sample size ratio (*q*: *N*)	
1: 5 to 1:10	Bentler and Chou (1987)
1:10	Jackson (2003)
1: 5 to 1:20	Kline (2015)

Overall, there are many steps in the questionnaire development, which require samples, as illustrated in **Figure 1**. Option one is generating an item pool where samples for interviews range from one to 50 and samples for the Delphi technique range from 10 to 100. Option two consists of testing content validity, where samples range between two and 20 experts. Option three is pretesting, and this ranges from 15-30 subjects. Option four is construct validity wherein factor analyses ranges from 50 to >1000.

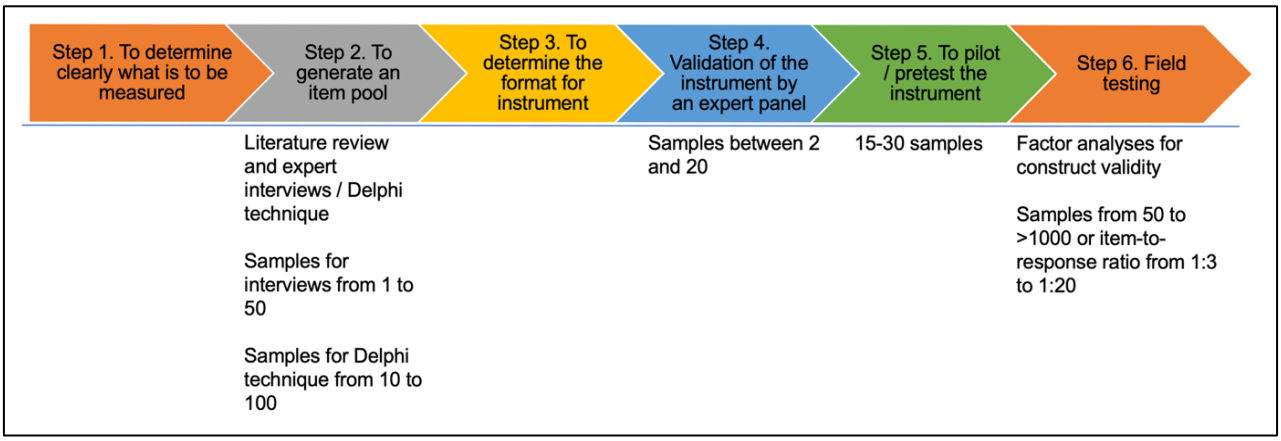

Figure 1 Instrument development steps requiring samples

Sample Size for Questionnaire Adaptation

Questionnaire adaptation is common in nursing research, but many studies lack information and transparency regarding why and how they adapt the questionnaire (Sullivan, 2011; Sousa et al., 2017). This lack of transparency may compromise the validity and reliability of the adapted questionnaire.

Questionnaire adaptation can be described in multiple ways: questionnaire translation; questionnaire modification by adding or removing items; and questionnaire adaptation. However, little changes carry significant implications for the overall questionnaire. These three strategies may or may not be conducive to construct validity with EFA/CFA. If the EFA/CFA is needed, additional samples are required according to the recommended sample sizes mentioned in the questionnaire development section.

In the case of instrument translation, such as from English to the Indonesian language, construct validity with factor analyses may, or may not, be needed if the researchers can ensure an accurate translation process to prevent meaning shifts and appropriate cultural adaptations. Each step of the translation, such as the use of the forward backward translation process and translation from experts, should be explained clearly. Otherwise, construct validity is needed if the translation is questionable. Mostly, the translation process occurs with content validity testing.

Questionnaire modification occurs when the researchers remove and/or add items, and in this case, construct validity is necessary. Adding and removing just one or two items may change the whole construct, and therefore, the meaning of the questionnaire, the factor structures, or latent variables may be shifted. Researchers should be meticulous in modifying the existing questionnaire, and a clear description should be made to provide a rationale.

Questionnaire adaptation, such as changing the setting, location, subject, or paraphrasing, may or may not require EFA or CFA. For example, if researchers only change the word of the location from “hospital” to “healthcare center” in the questionnaire, meaning shift may not occur. This is similar to paraphrasing, such as from “I feel anxious in this hospital” to “This hospital makes me feel anxious,” and there is no meaning shift identified. Because there is no meaning shift, there is no need for construct validity, however, content validity may be needed. When researchers change “anxious” to “worry/fear,” or change the subject from “I” to “they,” the meaning, while similar, is changed and construct validity testing is necessary. Thus, every detail in the questionnaire items that have been changed should be described clearly.

Conclusion

The appropriate sample size for questionnaire development and validation should be evaluated on an individual basis. Although general rules, item-to-response ratios, and parameter-to-sample ratios for factor analyses are expressed in sample size community norms, critical thinking is needed to consider the factors or variables that may influence sample size sufficiency, especially related to strong data, saturation, and other parameters pertaining to the specifics of the particular project.

It is also suggested that researchers not necessarily use 30 subjects for all validation scenarios, and it is recommended that the number in the instrument be carefully considered. Fifty responses are also not recommended for nursing research for a questionnaire, but it may be appropriate for obscure or difficult samples or chemical measurement. In any sample, it is paramount for researchers to provide a transparent presentation and explanation of such evidence-based judgment and rationale to ensure the appropriate sample size is established.

Declaration of Conflicting Interests
All authors declared that there is no conflict of interest.

Funding
None.

Authors’ Contributions
All authors contributed equally to this study.

Authors’ Biographies
Joko Gunawan, S.Kep.Ners, PhD is Director of Belitung Raya Foundation and Managing Editor of Belitung Nursing Journal, Bangka Belitung, Indonesia. He is also a Postdoctoral Researcher at the Faculty of Nursing, Chulalongkorn University, Bangkok, Thailand.
Colleen Marzilli, PhD, DNP, MBA, RN-BC, CCM, PHNA-BC, CNE, NEA-BC, FNAP is Associate Professor at the University of Texas at Tyler, USA. She is also on the Editorial Advisory Board of Belitung Nursing Journal.
Yupin Aungsuroch, PhD, RN is Associate Professor and Director of PhD Program at the Faculty of Nursing, Chulalongkorn University, Bangkok, Thailand. She is also an Editor-in-Chief of Belitung Nursing Journal.

References
Akins, R. B., Tolson, H., & Cole, B. R. (2005). Stability of response characteristics of a Delphi panel: Application of bootstrap data expansion. *BMC Medical Research Methodology, 5*(1), 1-12. https://doi.org/10.1186/1471-2288-5-37

Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement, 36*(4), 879-883. https://doi.org/10.1177%2F001316447603600410

Armstrong, T. S., Cohen, M. Z., Eriksen, L., & Cleeland, C. (2005). Content validity of self-report measurement instruments: An illustration from the development of the Brain Tumor Module of the MD Anderson Symptom Inventory. *Oncology Nursing Forum, 32*, 669-676.

Barrett, P. T., & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study and Group Behavior, 1*(1), 23-33.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117. https://doi.org/10.1177%2F0049124187016001004

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Burns, N., & Grove, S. K. (2005). The practice of nursing research: Conduct, critique and utilization. United States: Elsevier/Saunders.

Cattell, R. (1978). *The scientific use of factor analysis*. New York: Plenum.

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187-203). Washington, D.C: American Psychological Association.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*(5), 754-761.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DeVellis, R. F. (1991). *Scale development: theory and applications*. California: Sage publications.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *The British Journal of Psychiatry, 126*(3), 237-240. https://doi.org/10.1192/bjp.126.3.237

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265-275. https://psycnet.apa.org/doi/10.1037/0033-2909.103.2.265

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: mcGraw-Hill.

Hair, J. F., Black, B., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). London: Pearson.

Hatcher, L. (1994). *A step-by-step approach to using the SAS® system for factor analysis and structural equation modeling*. Cary, N.C: SAS Institutte, Inc.

Hutcheson, G., & N., S. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage Publication.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis. *Structural Equation Modeling, 10*(1), 128-141. https://doi.org/10.1207/S15328007SEM1001_6

Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guilford publications.

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE life sciences education, 18*(1), rm1-rm1. https://dx.doi.org/10.1187%2Fcbe.18-04-0064

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(08), 2207. https://doi.org/10.4236/psych.2018.98126

Lavrakas, P. J. (2008). Questionnaire *Encyclopedia of Survey Research Methods* (Vol. 1-10). Thousand Oaks, California: Sage Publications, Inc.

Morse, J. M. (2000). Determining sample size. *Qualitative Health Research, 10*(1), 3-5. https://doi.org/10.1177%2F104973200129118183

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159-168. https://doi.org/10.1207/s15327574ijt0502_4

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Rummel, R. J. (1988). *Applied factor analysis*. United States: Northwestern University Press.

Sousa, V. E. C., Matson, J., & Dunn Lopez, K. (2017). Questionnaire adapting: Little changes mean a lot. *Western Journal of Nursing Research, 39*(9), 1289-1300. https://doi.org/10.1177%2F0193945916678212

Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education, 3*, 119-120. https://doi.org/10.4300/JGME-D-11-00075.1

Watson, R., & Thompson, D. R. (2006). Use of factor analysis in Journal of Advanced Nursing: Literature review. *Journal of Advanced Nursing, 55*(3), 330-341. https://doi.org/10.1111/j.1365-2648.2006.03915.x

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), 165-178. https://dx.doi.org/10.15171%2Fjcs.2015.017

Cite this article as: Gunawan, J., Marzilli, C., & Aungsuroch, Y. (2021). Establishing appropriate sample size for developing and validating a questionnaire in nursing research. *Belitung Nursing Journal, 7*(5), 356-360. https://doi.org/10.33546/bnj.1927