EDUCATIONAL RESOURCES
Volume 13 Issue 1 2021
DOI: 10.21315/eimj2021.13.1.10
ARTICLE INFO
Submitted: 15-11-2020
Accepted: 15-02-2021
Online: 31-03-2021

ABC of Questionnaire Development and Validation for Survey Research

Muhamad Saiful Bahri Yusoff[1], Wan Nor Arifin[2], Siti Nurma Hanim Hadie[3]

[1]*Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA*

[2]*Biostatistics and Research Methodology Unit, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA*

[3]*Department of Anatomy, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA*

To cite this article: Yusoff MSB, Arifin WN, Hadie SNH. ABC of questionnaire development and validation for survey research. Education in Medicine Journal. 2021;13(1):97–108. https://doi.org/10.21315/eimj2021.13.1.10

To link to this article: https://doi.org/10.21315/eimj2021.13.1.10

ABSTRACT

Survey-based studies are ubiquitous in medical, social, economic, psychological and behavioural research, where questionnaires are often used as the main research tool to collect various information from respondents. Given the importance of questionnaires to research, ensuring the validity of the questionnaires is critical to producing high quality survey research. Therefore, this article describes a step-by-step systematic approach to questionnaire development and validation for research purposes.

Keywords: *Questionnaire development, Survey research, Validation, Psychometric, Validity*

CORRESPONDING AUTHOR
Muhamad Saiful Bahri Yusoff, Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia | Email: msaiful_bahri@usm.my

INTRODUCTION

Survey-based studies are abundant in medical, social, economic, psychological and behavioural research (1). In these studies, questionnaires are commonly used as a research tool to measure various information from the participants (1–2). Therefore, the use of a well-designed questionnaire is crucial in determining the quality and scientific merit of any survey-based research. Designing a questionnaire requires proper planning to ensure that relevant questions and items are considered in a way that reflects specific constructs to be measured in a research (3). Once the questionnaire has been designed, a crucial step is to perform a questionnaire validation to ensure quality responses and results (4). A rigorous systematic approach must be followed to design, develop and validate a questionnaire. This article outlines seven key steps for developing and validating questionnaires based on the best practices and our research experience. The first four steps focus on the design and development of a questionnaire, while the last three steps deal with the validation aspect of a questionnaire. Given the complexity of the questionnaire development and validation process with several examples to facilitate discussion of the key steps.

QUESTIONNAIRE DESIGN AND DEVELOPMENT

Step 1: Set Clear Aims

The first step in designing and developing a questionnaire is to set clear aims and goals for developing the questionnaire. The following are several questions to help researchers in setting the aims and goals. Answering these questions is pertinent for a successful questionnaire development.

a. What precisely will this questionnaire measure?

   For example, Medical Student Stressor Questionnaire (MSSQ) was developed to measure sources of stress among medical students (5) and Anatomy Education Environment Measurement Inventory (AEEMI) was developed to measure learning experience that influence medical students' motivation to learn anatomy subject, thus affecting their attitudes, values and behaviours towards anatomy-related learning tasks (6). It is essential to have a clear end in mind about the attributes (i.e., concepts, characteristic or features of someone or something) to be measured before developing a questionnaire.

b. Who is the intended target group?

   Knowing the specific target group (e.g., university students, diabetic patients, health workers) who will respond to the questionnaire is important to ensure the validity of the questionnaire. For instance, the intended target group for both MSSQ and AEEMI were medical students.

c. Why it needs to be developed?

   Defining a clear reason for developing a questionnaire for research is critical to avoid redundancy. For instance, prior to the development of MSSQ, there are many questionnaires that measure stress level, but none was developed to measure the sources of stress among medical students. Considering this gap, MSSQ was developed and validated to measure the sources of stress among medical students (5). This would ensure the uniqueness of a questionnaire, as the researchers do not reinvent a tool that resembles an existing inventory and hence avoid wasting time and resources.

d. How will it contribute to practices in the field?

   Stating clearly the expected contributions of the questionnaire under development is important to ensure its relevance to the current practices in the field. This will also confirm that the effort is not to "reinvent the wheel". Taking MSSQ as an example, the inventory could become a universal tool to identify sources of stress among medical students and encourage medical educators around the globe to evaluate the potential sources of stress among their students, therefore early interventions could be planned to alleviate the stressors (5).

Step 2: Define Attributes

Once the what, who, why and how questions are clearly answered, the next step is to define the attributes representing the intended outcomes to be measured by the questionnaire. In general, an attribute refers to the quality, feature or characteristic of a subject, either a person or an object. In the context of a questionnaire, it is conceptualised as the construct, domain or factor to be measured. Having a good understanding of relevant theories and literature can help us to come out with a suitable definition for each attribute. Having clear definitions is important to help us identify relevant items to represent the attributes. Clear descriptions of the intended attributes will help in defining the nature of the attributes and their items. The following are the recommended strategies for defining the attributes:

a. Conduct a literature review

A complete and thorough literature review is important in order to gain a solid basic understanding of the attributes and their elaboration from previous research, to identify other existing attributes and to acquire items related to the attributes. A literature review provides a framework for what the questionnaire could look like and how it would differ from the existing questionnaires.

b. Conduct interviews and/or focus groups

These strategies allow researchers to gain information from the target population on how they conceptualise and describe the attributes of interest. Interestingly, these strategies make the concept of attributes understandable in the language commonly used without violating the theoretical sense of the scholars. Failure to clarify exact attributes suggests an incoherent and invalid assessment. An example of detailed descriptions of attributes measured by MSSQ is summarised in Table 1.

Table 1: Identified MSSQ attributes and the description of each attribute

Identified attribute	Description
Academic related stressors	The literature showed that the main stressors were tests and examinations, time pressure and getting behind in work as well as conflicting demands, not getting work done within time planned and heavy workload. In addition, students who are perfectionists (high self-expectations) are at greater risk for psychological distress due to high self-expectation to do well in examinations. All of these stressors were related to the academic matter.
Intrapersonal and interpersonal related stressors	The literature showed that intrapersonal conflict, interpersonal interaction and relationship were stressors for medical students; for example poor motivation to learn, conflict with other students, teachers and personnel. These stressors were related to intrapersonal and interpersonal relationship issues.
Teaching and learning related stressors	The literature showed that dissatisfaction with the quality of education, lectures, guidance and feedback from teachers, and recognition of work done as well as uncertainty of what is expected from the students were perceived by medical students as stressors. All of these stressors were generally related to the teaching and learning process.
Social related stressors	The literature showed that the level of dissatisfaction in social activities was associated with psychological distress among medical students. A significant relationship between emotional disorder and medical students' relationship with their family and friends, facing illness or death of patients and inability to provide appropriate answer to patients were stressors perceived by medical students. In addition, dissatisfaction with social activities was associated with emotional disorder. These stressors were generally related to the social relationship between the students and other people such as family and patients.
Drive and desire related stressors	The literature showed that political and family pressures as well as fear of choosing wrong career and unwillingness to study medicine were recognised as stressors in medical students. All the stressors were related to the motivation of the students to learn medicine.

Step 3: Write a Plan

Planning is an important step in ensuring that the development and validation process runs smoothly. The following are several areas that must be considered during the planning stage:

a. Test content

Test content can be determined by (i) using a grid-style blueprint to determine the content areas and how these are potentially manifested by the subjects, (ii) bringing in a small group of stakeholders to brainstorm the various angles and facets of an attribute, or (iii) including subjects who are at the extremes of the attribute so that you can identify item content that reflects the entire spectrum.

b. Target population

It should be clearly defined to ensure a proper conclusion.

c. Number and types of relevant items

All relevant aspects of each attribute need to be identified and constructed to form the items. For a relatively simple questionnaire, a rule of thumb of 10 items per attributes is applicable during the developmental stage.

d. Administration instructions

The instructions should be clearly designed, especially for a self-reporting questionnaire. Clear instructions will guide the respondents about what and how to complete the questionnaire, therefore, strengthening the response process validity.

e. Estimate the completion time

This will depend on the types of questionnaire (e.g., close-ended, open-ended, scaling type, matrix type and ranking type questionnaires). It is advisable to develop a questionnaire that takes less time to complete (less than 15 minutes if possible). The shorter the completion time, the better the response rate.

f. How scores should be calculated and interpreted

The easiest way to get a score in a questionnaire is by calculating the sum of the responses. However, a negatively worded item requires a reverse scoring before calculating the sum. It is also important to provide a clear interpretation of the scores as a guide to the users and to ensure the consistency of their meaning across studies.

Step 4: Develop and Write Items

The final step in questionnaire development is to write items and to design the most appropriate response format, which will be presented in a form of an answer sheet. There are several types of response format such as semantic and Likert scale. For the Likert-type scale response anchors, we recommend researchers to refer to the article written by Vagias (7). Write each item clearly, keep the item as short as possible and avoid double negatives. In order to reduce the response bias, which is the tendency of respondent to give the same answer to every item, some items should be reverse-phrased (i.e., negative items). Nevertheless, these items must be reverse-scored during the analysis. The item layout items should be simple and straightforward, so as to allow respondents to easily link each item to its respective response options. This is important to ensure the response process validity of the questionnaire.

QUESTIONNAIRE VALIDATION

Step 5: Select Items

The following strategies are recommended for selecting items (2, 8–13):

Conduct content validation

a. It is to assess the items that are relevant and representative of the construct of interest (9).

b. Content validity index (CVI) – items with a CVI value of at least 0.80 should be kept in the questionnaire. Readers are recommended to refer the article by Yusoff (8) for detailed steps in conducting content validation and calculating the CVI .

Conduct cognitive interview and/or response process validation

a. It is to ensure that respondents interpret the items as intended by the survey designer (2, 9).

b. Cognitive interviewing – readers are recommended to refer to the article written by Beatty and Willis (12) on the practice of cognitive interviewing.

Conduct field pre-testing (pilot study)

a. This is conducted on a representative sample to evaluate the face validity, administration process, data entry preparation and descriptive statistics (2, 11, 14). A minimum number of 25 to 75 respondents is recommended at this stage (14). In general, we suggest that a sample of 30 respondents should be sufficient, which is a typical size for a pilot study in medical research (15–16).

b. Face validity index – face validity evaluation allows evaluation of item clarity and comprehension (9). Items with values of at least 0.80 should be kept in the questionnaire. Researchers are recommended to refer to the article by Yusoff (9) for detailed steps in conducting response process validation and calculating face validity index.

c. Form administration process (14) – it is important to evaluate how long it takes for a respondent to complete the questionnaire. For interviewer-administered questionnaire, rater training and reliability must be guaranteed to allow a standardised way of interviewing the respondents. In addition, the flow/process of administering the questionnaire (the technical details, for example how to distribute and collect the questionnaire) and the logistical aspects of administration must be taken care of.

d. Data entry preparation – the pre-testing stage also allows fine-tuning of the data entry form preparation, data entry procedure and data coding (14). Issues with the data entry should be anticipated earlier and modifications can be done to the questionnaire for the subsequent stages of the study.

e. Descriptive statistics

   i. Calculate the percentage of responses per scale rating, minimum and maximum rating. These are obtained to ensure all range of the scale is utilised (e.g., 1 to 5).

   ii. Check the floor and ceiling effects. The effects are present when more than 15% of responses are at the lowest and/or the highest ends of the scale (17). For example, responses for an item that are clustered at the highest end might need additional ratings for the scale at the end of the scale, while it might also point to the problem with the item statement (e.g., item statement that will result in most participants answer "strongly agree" with then item). In addition, take note of the ratings that are not utilised/under utilised, as these point to either the problem with the rating (not relevant or applicable) or the item statements.

iii. Calculate mean (standard deviation) or median (interquartile range). These provide basic statistical estimates for the items. Since the sample size will be small in a pilot study, the statistics only give rough estimates about the distribution of the responses per item, and the values are best verified in a larger sample size during the psychometric evaluation.

Conduct psychometric evaluation of the internal structure

a. It is to check the internal structure of the questionnaire, which is concerned with item-item, item-construct and construct-construct interrelationships. This is generally obtained by factor analysis and reliability (18).

b. Factor analysis – the aim of factor analysis is to find the number and nature of common factors that explain the pattern of correlations between items (19–20). The common factors are unobserved (latent) (19–21), thus factor analysis allows researchers to explore (exploratory factor analysis, EFA) or confirm (confirmatory factor analysis, CFA) the presence of underlying hypothesised constructs, measured by the corresponding items.

   i. Factor loading – for EFA, a factor loading between 0.3 to 0.4 is minimally acceptable (20, 22). For CFA, it is recommended to aim for a factor loading of more than 0.5 (20). However, to the authors' experience, a lower cutoff value of more than 0.3 can be accepted if it is deemed very important to keep the items after considering the content of the item statements, provided the model fit and reliability remain good.

   ii. Factor-factor correlation – factors must be distinct/discriminant from each other. This is indicated by a correlation coefficient (*r*) value of less than 0.85 (19).

   iii. Sample size – the sample size for CFA depends on factor loadings, number of factors and number of items (23). In general, for EFA we are comfortable with a sample size of at least 50 respondents (23) or respondents-to-items ratio of 5:1 (10). In general, for CFA we suggest a sample size of at least 200 respondents (24). Researchers may determine the sample size for CFA by Monte Carlo simulation (25) or sample size formulas for fit indices of choice (26). One of the authors developed a web-based sample size calculator based on expected comparative fit index (CFI) and root mean square of error approximation (RMSEA) (27).

   iv. The goodness of fit indices – In CFA, the latent constructs or proposed models are considered fit if all the goodness of fit indices achieved the minimal requirement (28). Table 2 summarises commonly used goodness of fit indices to signify model fit (28).

Table 2: Goodness of fit indices used to signify model fit

Category	Name of index	Level of acceptance
Absolute fit[1]	RMSEA	less than 0.08
	Goodness of fit index (GFI)	more than 0.9
Incremental fit[2]	CFI	more than 0.9
	Tucker-Lewis index (TLI)	more than 0.9
	Normed fit index (NFI)	more than 0.9
Parsimonious fit[3]	Chi-square/degree of freedom (χ^2/df)	less than 5

Note: [1]Absolute fit: Measures overall goodness of fit for both the structural and measurement models collectively. This type of measure does not make any comparison to a specified null model (incremental fit measure) or adjust for the number of parameters in the estimated model (parsimonious fit measure). [2]Incremental fit: Measures goodness of fit that compares the current model to a specified "null" (independence) model to determine the degree of improvement over the null model. [3]Parsimonious fit: Measures goodness of fit representing the degree of model fit per estimated coefficient. This measure attempts to correct for any "overfitting" of the model and evaluates the parsimony of the model compared to the goodness of fit.

c. Reliability – this is indicated by high value of internal consistency reliability coefficient, commonly measured by Cronbach's alpha coefficient (α).

   i. Cronbach's alpha (α) value of more than 0.7 signifies an acceptable level of internal consistency reliability (29). The sample size for determination of Cronbach's alpha can be calculated based on the acceptable and expected values of the coefficient (30–31).

   ii. Corrected item-total correlation – ideally a value of more than 0.5 shows a good correlation between the item with the total without the item (20). However, the minimum acceptable value is not less than 0.30 (32).

   iii. Cronbach's alpha if item deleted – if deleting the item reduces the Cronbach's alpha, this signifies the item should be retained. If deleting the item increases the Cronbach's alpha, this signifies the item should be removed.

Conduct additional psychometric evaluation

It is done after the construct validity by internal structure has been established. It explores other sources of evidence to support the validity of the questionnaire (18). One of the important sources of evidence to highlight is the relationship with other variables/measures. To establish the relationship with other measures (2, 10–11), researchers can administer other related (convergent) or unrelated (discriminant) measures together with the questionnaire tested during data collection in Step 5 (conduct psychometric evaluation of the internal structure) above. To establish the relationship with other variables, it is recommended to do so in another sample. A different sample size determination is required, depending on the planned analysis (e.g., correlation, group comparison, regression). For a practical example of the psychometric evaluation and the sources of evidence of validity, readers may refer to the systematic review by Yusoff (5).

Step 6: Standardisation of Scores

Standardisation of scores is important to allow users to interpret scores by creating norms or scoring systems. Norms or scoring systems are useful for interpreting scores by comparing the score groups. This is essential for a questionnaire being developed for research purposes. The following methods are recommended for standardising scores:

a. Standardised score based on mean and standard deviation (z-score): To calculate a standardised score (z-score), subtract the mean from the raw score and divide by the standard deviation (33).

$$z\text{-score} = (\text{m} - \mu) / \text{sd}$$

m = observed score
μ = mean score
sd = standard deviation

To convert a z-score to an observed score, multiply the z-score by the standard deviation and add the mean score:

$$\text{m} = (\text{sd} \times z\text{-score}) + \mu$$

b. Percentile rank of scores: Based on percentiles derived from a normative sample, the following classification ranges and their corresponding percentile rank ranges are commonly used; 0th to 24th percentile represents low, 25th to 75th percentile represents average (i.e., 25th to 50th percentile represent low average, 51th to 75th percentile represent high average), 76th to 100th percentile represent high. The following is the common percentile formula used nowadays (34):

$$P = (i - 0.5) / n \times 100)$$

P = percentile
i = rank of data
n = number of data

c. Ebel's method: Three levels of scores can be established based on the Ebel's method that classify scores by the highest 27% (high), the lowest 27% (low) and the in-between highest-lowest (average) (35).

Step 7: Final Preparation

Any final amendments can now be made. It is necessary to ensure that the instructions and answer/response sheets are satisfactory, the scoring procedures are working effectively and the scores can be interpreted intuitively. In the case of personality questionnaires, some form of profile sheet is often needed to illustrate the different scores on the scales.

CONCLUSION

This article describes a step-by-step guide to developing and validating questionnaires for survey research. The summary of the seven steps for developing and validating a research questionnaire is shown in Table 3. The seven-step guideline is intended to systematically guide researchers in developing a valid questionnaire for their research.

Table 3: The worksheet for developing and validating questionnaire for research

Key steps	Description
Step 1: Set clear aims	We begin by providing specific answers to the following questions. Without clear answers to these questions, your measure may not be useful: a. What precisely will this questionnaire measure? b. Who is the intended target group? c. Why it needs to be developed? d. How will it contribute to practice in the field?
Step 2: Define attributes	Having a good understanding of relevant psychological theories and literature can help identify what items should or should not be included. The following are the recommended strategies: a. Conduct a literature review – help to have a sound basic understanding of the attribute and other research involving it. You might also be able to identify other existing measures and to consider, therefore, what kinds of items are needed, what your questionnaire might look like and how it will differ from them. b. Conduct interviews and/or focus groups – to learn how the population of interest conceptualises and describes the construct of interest. Both strategies ensure that the conceptualisation of the construct makes theoretical sense to scholars in the field and uses language that the population of interest understands. Failure to clarify exactly what is measured could mean that you end up with an assessment which is confusing.
Step 3: Write a plan	Think about how we will go about determining the following: a. Test content – (i) use a grid-style blueprint to determine the content areas and how these are potentially manifested by the subjects, (ii) get a small group of people to brainstorm as many facets of the construct as possible, or (iii) include people who might be at the extremes of your construct so that you can identify item content which reflects the entire spectrum. b. Target population – it should be defined clearly. c. The types of items needed and their number – we need to reflect all relevant aspects of the attribute. For a relatively simple measure, aim for at least 10 items per attribute. d. Administration instructions – it must be clearly developed, especially for a self-reporting questionnaire. e. Completion time – this depends on the kind of measure. f. How scores should be calculated and interpreted – the simplest is to sum the responses. If some items are negatively worded, reverse the scores before calculating the sum.
Step 4: Develop and write items	a. Write items and consider the most appropriate response format. b. Write each item clearly, keep the item as short as possible and avoid double negatives. c. To reduce response bias, reverse-phrase some of the items. Reverse-score before analysis. d. Item layout should be simple and straightforward, that allows respondents to link each item to its respective response options.

(*continued on next page*)

Table 3: (*continued*)

Key steps	Description
Step 5: Select items	The following strategies are recommended for selecting items: **Conduct content validation** – to assess the items that are relevant and representative of the construct of interest **Conduct cognitive interview and/or response process validation** – to ensure that respondents interpret the items as intended by the survey designer. **Conduct pilot testing** - to evaluate the face validity, administration process, data entry preparation and descriptive statistics. a. Face validity index. b. Form administration process. c. Data entry preparation. d. Descriptive statistics i. Percentage of responses, minimum and maximum rating. ii. Floor and ceiling effects. iii. Mean (standard deviation) or median (interquartile range). **Conduct psychometric evaluation of the internal structure** – to check the internal structure of the questionnaire, which is concerned with item-item, item-construct and construct-construct interrelationships. a. Factor analysis – to find the number and nature of common factors that explain the pattern of correlations between items. b. Reliability – high level of internal consistency. **Conduct additional psychometric evaluation** – it is done to explore other sources of evidence to support the validity of the questionnaire.
Step 6: Standardisation of scores	Standardisation of scores is important to allow users to interpret scores by creating norms or scoring systems. Norms or scoring systems are useful for interpreting scores by comparing the score groups. This is essential for a questionnaire being developed for research purposes. The recommended methods for standardising scores are: *z*-score, percentile rank of scores and Ebel's method.
Step 7: Final preparation	Any final amendments can now be made. It is necessary to ensure that the instructions and answer/response sheets are satisfactory, the scoring procedures are working effectively and the scores can be interpreted intuitively. In the case of personality questionnaires, some form of profile sheet is often needed to illustrate the different scores on the scales.

REFERENCES

1. Saris WE, Gallhofer IN. Design, evaluation, and analysis of questionnaires for survey research. New Jersey: John Wiley and Sons; 2007. https://doi.org/10.1002/9780470165195

2. Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE guide no. 87. Med Teach. 2014;36(6):463–74. https://doi.org/10.3109/0142159X.2014.889814

3. Acharya B. Questionnaire design. Central Department of Population Studies. Working paper to be discussed at Training–cum-Workshop in Research Methodology; 2010 June 5; Pulchok, Lalitpur, Nepal. 2010.

4. Williams A. How to…write and analyse a questionnaire. J Orthod. 2003;30(3):245–52. https://doi.org/10.1093/ortho/30.3.245

5. Yusoff MSB. A systematic review on validity evidence of medical student stressor questionnaire. Education in Medicine Journal. 2017;9(1):1–16. https://doi.org/10.21315/eimj2017.9.1.1

6. Hadie SNH, Hassan A, Ismail ZIM, Asari MA, Khan AA, Kasim F, et al. Anatomy education environment measurement inventory: a valid tool to measure the anatomy learning environment. Anat Sci Educ. 2017;10(5):423–32. https://doi.org/10.1002/ase.1683

7. Vagias WM. Likert-type scale response anchors. Clemson, SC: Clemson University; 2006.

8. Yusoff MSB. ABC of content validation and content validity index calculation. Education in Medicine Journal. 2019;11(2):49–54. https://doi.org/10.21315/eimj2019.11.2.6

9. Yusoff MSB. ABC of response process validation and face validity index calculation. Education in Medicine Journal. 2019;11(3):55–61. https://doi.org/10.21315/eimj2019.11.3.6

10. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research, and Evaluation. 2005;10:art.7. https://doi.org/10.7275/jyj1-4868.

11. Coaley K. An introduction to psychological assessment and psychometrics. London: Sage Publications; 2010.

12. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. Public Opinion Quarterly. 2007;71(2):287–311. https://doi.org/10.1093/poq/nfm006

13. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas. 2013;76(6):913–34. https://doi.org/10.1177/0013164413495237.

14. McDonald J, Burnett N, Corodano V, Johnson R. Questionnaire design. Georgia: Division of Reproductive Health; 2003.

15. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract. 2004;10(2):307–12. https://doi.org/10.1111/j..2002.384.doc.x.

16. Browne RH. On the use of a pilot sample for sample size determination. Stat Med. 1995;14(17):1933–40. https://doi.org/10.1002/sim.4780141709.

17. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

18. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166.e7–16. https://doi.org/10.1016/j.amjmed.2005.10.036

19. Brown T. Confirmatory factor analysis for applied research. 2nd ed. New York, US: Guilford Publications; 2015.

20. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. London, UK: Cengage Learning EMEA; 2010.

21. Bartholomew DJ, Steele F, Galbraith J, Moustaki I. Analysis of multivariate social science data. 2nd ed. New York, US: Chapman and Hall/CRC; 2008. https://doi.org/10.1201/b15114

22. Stevens JP. Applied multivariate statistics for the social sciences. New York, US: Routledge; 2012. https://doi.org/10.4324/9780203843130

23. de Winter JCF, Dodou D, Wieringa PA. Exploratory factor analysis with small sample sizes. Multivariate Behav Res. 2009;4(2):147–81. https://doi.org/10.1080/00273170902794206

24. Klein RB. Principles and practice of structural equation modelling. 4th ed. New York, US: Guilford Press; 2016.

25. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling: A Multidisciplinary Journal. 2002;9(4):599–620. https://doi.org/10.1207/S15328007SEM0904_8

26. Kim KH. The relation among fit indexes, power, and sample size in structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal. 2005;12(3):368–90. https://doi.org/10.1207/s15328007sem1203_2

27. Arifin WN. Sample size calculator [Internet]. c2017–2021 [cited 2021 Jan 26]. Available from: http://wnarifin.github.io

28. Schreiber JB, Stage FK, King J, Nora A, Barlow EA. Reporting structural equation modeling and confirmatory factor analysis results: a review. Journal of Educational Research. 2006;99(6):323–38. https://doi.org/10.3200/JOER.99.6.323-338

29. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ. 1997;314:572. https://doi.org/10.1136/bmj.314.7080.572

30. Arifin WN. A web-based sample size calculator for reliability studies. Education in Medicine Journal. 2018;10(3):67–76. https://doi.org/10.21315/eimj2018.10.3.8

31. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. J Educ Behav Stat. 2002;27(4). https://doi.org/10.3102/10769986027004335

32. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.

33. Siegle D. Standardized scores [Internet]. Neag School of Education, University of Connecticut. [cited 2021 Jan 26]. Available from: https://researchbasics.education.uconn.edu/standardized-scores/#:~:text=To calculate a z-score,15 minus 10 equals 5.

34. Bornmann L, Leydesdorff L, Mutz R. The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. Journal of Informetrics. 2013;7(1):158–65. https://doi.org/10.1016/j.joi.2012.10.001

35. Ebel R. Measuring educational achievement. United State: Prentice Hall; 1965. p.348–9.