Methodological Issues in Questionnaire Design

Song, Youngshin[1] · Son, Youn-Jung[2] · Oh, Doonam[3]

[1]College of Nursing, Chungnam National University, Daejeon
[2]Department of Nursing, Soonchunhyang University, Cheonan
[3]Department of Nursing, Hoseo University, Asan, Korea

Purpose: The process of designing a questionnaire is complicated. Many questionnaires on nursing phenomena have been developed and used by nursing researchers. The purpose of this paper was to discuss questionnaire design and factors that should be considered when using existing scales. Methods: Methodological issues were discussed, such as factors in the design of questions, steps in developing questionnaires, wording and formatting methods for items, and administrations methods. How to use existing scales, how to facilitate cultural adaptation, and how to prevent socially desirable responding were discussed. Moreover, the triangulation method in questionnaire development was introduced. Results: Steps were recommended for designing questions such as appropriately operationalizing key concepts for the target population, clearly formatting response options, generating items and confirming final items through face or content validity, sufficiently piloting the questionnaire using item analysis, demonstrating reliability and validity, finalizing the scale, and training the administrator. Psychometric properties and cultural equivalence should be evaluated prior to administration when using an existing questionnaire and performing cultural adaptation. Conclusion: In the context of well-defined nursing phenomena, logical and systematic methods will contribute to the development of simple and precise questionnaires.

Key words: Questionnaire, Design, Cultural, Nursing research

INTRODUCTION

Psycho-social-behavioral concepts are commonly measured by nursing researchers using questionnaires.

Questionnaires are useful and easy to administer to collect data from participants in studies[1]. It is absolutely important for a researcher to be aware of the importance of a well-designed questionnaire and whether it measures what it is intended to measure. Therefore, the use of a valid and reliable tool to measure the properties of psych-social-behavioral concepts is an essential part of well-designed studies. Consideration should be given accordingly on whether the questionnaire will measure quantitative or qualitative data, and what would be its mode of administration.

With respect to questionnaire design, we present a definition of measurement, the guiding basic principles of measurement, and broad overview to the readers on questionnaires. Moreover, we discuss several important issues in the enhancement of the reliability and validity of measurement.

CONSIDERATIONS PRIOR TO QUESTIONNAIRE DESIGN

Most respondents have the tendency to respond to questionnaires without considering how missing responses will be analyzed, how they will contribute to answering research questions, and how researchers will account for questionnaires that are not

returned by mail. Most researchers experience issues related to non-response when self-report questionnaires are used. The literature has offered suggestions on how to avoid those problems and how to develop questionnaires to measure psychological constructs more concisely. Frary[2] presented considerations prior to questionnaire design such as the investigator must define precisely the information desired to write as few questions as possible to obtain it, and in a second step is to obtain feedback from a small but representative sample of potential responders. Frary[2] recommended that a field trial might be desirable or necessary if there is substantial uncertainty in the following areas: a) Response rate: if a field trial of a mailed questionnaire yields an unsatisfactory response rate, design changes or different data gathering procedures must be undertaken. b) Question applicability: Even though approved by reviewers, some questions may prove redundant. For example, everyone or nearly everyone may be in the same answer category for some questions, thus making them unnecessary. c) Question performance: the field-trial response distributions for some questions may clearly indicate that they are defective.

STEPS IN DESIGNING MEASUREMENT TOOLS

The process and steps for developing a scale vary depending on what is being measured in a study. Stehr-Green et al.[3] summarized eight steps in creating a questionnaire for a successful epidemic study as follows: a) Identify the leading hypotheses about the source of the problem and b) the information needed to test the hypotheses, c) Identify the information needed for the logistics of the study and to examine confounding factors, d) Write the questions to collect this information, e) Organize the questions into questionnaire format, f) Test the questionnaire, g) Revise the questionnaire, and h) Train interviewers to administer the questionnaire. Colosi[1] recommended steps to developing an effective questionnaire when evaluating one's own program using a questionnaire. That is, the researcher should decide what kind of information to collect, and then review previous literature to obtain permission to use an existing questionnaire, or develop a new questionnaire. Then, the existing or newly developed questions should be modified or fit to the researcher's needs in a logical order. Finally, the researcher should re-read to clarify information for questions, or add specific instructions or transitions in

parentheses where applicable. At this point, the researcher (along with colleagues) should focus on the format of the questionnaire with attention to layout, readability, time demands on respondents, logic, and clarity of content. If necessary, the researcher can revise the instrument as needed based on feedback provided and prepare a protocol for implementing the questionnaire[1]. From a methodological perspective, Rattray and Jones[4] emphasized that a logical, systematic, and structured approach should be employed for questionnaire design, from item generation to psychometric evaluation. They, particularly, emphasized the importance of testing and pilot items, amendments based on item-analysis, principal components analysis, reliability, concurrent validity, confirmation using an independent data set, and revision of the measure[4]. Netemeyer et al.[5] introduced four steps for developing paper-and-pencil measures of social-psychological constructs. The first step is to choose the construct definition and content domain[5]. The second step involves generating and judging each item, and then designing and conducting research to develop and refine the scale. Lastly, the scale can be finalized[5].

We suggest strategies for designing questionnaires based on the various recommendations in the literature.

a) Appropriately operationalize the key concept for the target population.

b) Choose a clear response format.

c) Generate items and confirm final items using face or content validity.

d) Sufficiently pilot the questionnaire using item-analysis.

e) Demonstrate reliability and validity.

f) Finalize the scale and train the administrator.

DETAILS OF QUESTIONNAIRE DESIGN

Appropriate questionnaire design is essential to ensure valid responses to questions. The main purposes in designing questionnaires are commonly to obtain accurate relevant information and to maximize the response rate for the survey[6].

1. Order and wording of items

When generating the questionnaire, consider that the order of the items may play a considerable role in responses. Rattray and

Jones[4] recommended that controversial or emotive items should not be placed at the beginning of the questionnaire and demographic and/or clinical data may be presented at the end to keep respondents engaged.

Consideration should be given to the wording of questions, that is, technical jargon, slang, and abbreviations should be avoided[3]. The reading level of items should correspond to the level of education of respondents. Stehr-Green et al.[3] noted that each item should contain a single idea and double negatives should be avoided. Short and simple questions are generally recommended, because participants tend to have higher response rates and a higher proportion of completed answers for shorter items than for more complex items[6]. In particular, items such as leading questions, double-barreled questions, unclear and ambiguous questions, and invasive or personal questions should be avoided[1]. Leading questions can induce respondents to offer the researcher's preferred answers and double-barreled questions can confuse respondents. In addition, a mixture of both positively and negatively worded items are recommended to minimize the tendency for respondents to respond in the same way to items[4,6].

2. Formatting and arranging items

Response choices may include open-ended, fill-in-the-blank, and closed-ended formats[3]. Open-ended questions allow respondents to give answers based on their own perspectives for questions[1,3]. Open-ended items are useful when exploring the range of possible responses to a question; however, it is not easy to capture group information in this manner[1]. Thus, open-ended questions can be used as a preliminary method with a small sample to determine common themes in advance[2].

On the other hand, closed-ended designs can provide summary information and minimize bias against the less literate or articulate[6]. Closed-ended items are easy to administer and analyze. Closed formats include choice of categories, Likert-style scale (e.g., strongly agree, agree, cannot decide, disagree, strongly disagree), differential scales (e.g., extremely interesting to extremely dull, rated on a 10 point scale), checklists, and rankings[6].

Fill-in-the-blank format is similar to open-ended questions. Fill-in-the-blank questions are used when the response will be a relatively simple word or number[3]. This format can be used

when the question measures a simple respondent attribute (age, educational level), collects a date (birthdate, number of exposures), or quantifies something specific[3].

According to recommendations by Leung[6], there are several general rules for arranging such questions. The question should be ordered a) from general to particular, b) from easy to difficult, and c) from factual to abstract. Moreover, the items should start with closed-format questions, and questions relevant to the main subject.

3. Questionnaire administration

Self-administration is the most popular method of administering questionnaires in survey studies. Self-administered questionnaires can be collected via post, email, or electronically[6]. Self-administered questionnaires are easy to implement, cost-effective, and protect confidentiality. Moreover, they can be completed at the respondent's convenience and administered in a standard manner[6]. Interview-administered questionnaires can be conducted by telephone or face-to-face. Interview-administered questionnaires allow participation by illiterate people and clarification of ambiguity. The best method for administering questionnaires depends on who the respondents are. In any case, it is important to collect the right information, from the right population, at the right time, using the right method.

MEASUREMENT ISSUES FOR INCREASING RELIABILITY AND VALIDITY

1. Use of an existing questionnaire

Many researchers have focused on instrument development to measure health phenomena. As a result, appropriate instruments can be easily found for use in research and practice. Use of existing instruments may provide the advantage of cost-effectiveness and knowledge accumulation; however, instruments should be used in the same way that they were designed, to fit the situation in terms of place, time, and population[7].

When measuring a concept of interest, a preliminary search for an existing instrument is conducted. Likewise, searching for an existing instrument is the first step in defining the parameters and context of your concept. Waltz et al.[7] suggested several

tips for using databases when searching for existing instruments: a) search computerized database by using the name of keywords or instrument, b) generalized the search to specific area of interest, c) search for summary articles describing and evaluating the instruments used to measure a given concept, d) search journals that are devoted specially to measure, e) after identifying a publication in which relevant instrument are used, use citation indices to locate other publications that used them etc.

After identifying an instrument, it should be evaluated for adequacy in terms of its purpose and stated aims, measurement framework, conceptual basis, and psychometric properties. In particular, a psychometric evaluation should be performed before the existing instrument is chosen for use. Estimates of reliability, specificity, sensitivity, and validity based on psychometric testing ensure the appropriateness of the given instrument. In addition, whether an existing instrument corresponds to the specific population characteristics, place, and time for the intended setting should be considered[7].

If an existing instrument is identified, permission to use the instrument for a specific purpose should be obtained in writing from the developer or copyright holder[7]. This process is part of the legal and ethical responsibility of a user. If a given instrument requires modification, revised contents should be given to the developer. Moreover, the user has the responsibility to report and share results regarding the tool's properties, the nature of the sample, and the diversity of conditions[7].

2. Cross-cultural adaptations

Cross-cultural research collaboration to address global health issues is meaningful in terms of providing evidence for practice across cultures or nations. Cross-cultural measurement should be transferable across cultures, settings, and sites[7]. The terms cross-cultural and cross-national have sometimes been used interchangeably. However, cross-national research is always cross-cultural, but cross-cultural research may not always be cross-national[8,9]. When an existing tool is selected to measure cross-culturally, applicability is crucial. Waltz and colleagues[7] suggested that several queries should be addressed while selecting a measure. Evaluations regarding measurement items may include the following: a) items reflect cul-

turally relevant theoretical propositions that served as the basis for the measure's development, b) the type of measure is appropriate for the culture in terms of applicability, c) the scores can provide information that will help in decision making with respect to the phenomena of interest within the culture, d) the measure can be conducted for the study aims, e) the measure will be administered consistent with the intended conditions and settings, f) the results from measure are likely to be congruent with the intended setting philosophy, subjects, financial and administrative structure, g) the target population is similar to that in the culture, h) the time and resources including copyright permission to use, time required to administer and evaluate the measure as appropriate for the setting in which it will be applied[7] should be considered.

When measures are conducted across cultures, cultural equivalence and cultural bias have emerged as major issues. Cultural equivalence in measurement refers to the concept that scores obtained from a measure are similar when employed in different cultural populations[10]. Equivalence can be assessed by conducting a pretest with the populations of interest from various cultural backgrounds. Through this pretest process, similarities and differences in response patterns can be explored. Occasionally, culture-specific concepts such as depression may be eliminated in the translation process. Thus, cultural equivalence should be established before translation. If concepts differ in meaning across cultures, cultural bias may be revealed. According to Waltz and colleagues[7], three types of bias, construct, method, and item bias, may result because of poor item translation, inappropriate content of items, and unstandardized procedures. In particular, construct bias can be manifested when the following exists: a) differences in the appropriateness of content, b) inadequate sampling, c) underrepresentation of the construct, and d) incomplete overlap of the construct across cultures[10]. To avoid these biases, pretest procedures, adequate training of individuals to conduct the assessment in a culturally relevant manner, proper sampling, and accurate translation are commonly recommended[7,8,10].

As presented previously, accurate translation not only strengthens equivalence, but also helps to avoid bias. However, when linguistic differences such as the literal meaning of words between the source and target language exist, the process of translation is more difficult. In this case, researchers have generally employed

several translation strategies including the following: a) translation from the source to the target language and then from the target language back to the source language (back-translation), b) committee review of the translation and back-translation, c) pretesting for cultural equivalence, and d) confirming conceptual equivalence[7]. The translated version of the instrument should be tested in terms of its psychometric properties in the target group to identify evidence of reliability and validity. Many examples of evaluative and psychometric testing for various versions of one instrument in different languages can be found in the literature.

3. Social desirability

Social desirability is the extent to which individuals tend to project favorable images of themselves during social interactions[7]. That is, participants who are asked sensitive questions respond in a favorable manner to items with a socially desirable response[11]. For example, when the contents of a question invade a respondent's privacy, when an answer potentially causes risk to other parties, and when an answer may not be allowed as a social rule, the response is likely to be a socially desirable response. This type of responding can lead to various results through the interpretation of the related item, such as those measuring affective features[11]. To minimize socially desirable responses, several strategies have suggested. Knowles[12] reported that socially desirable responses can be increased when participants should answer the same general issue repeatedly across the items in single-dimension scale[13,14]. Therefore, item formats that generate more thought and polarized answers help to minimize socially desirable responses. Participants who respond in a socially desirable manner tend to believe that their information will not be kept confidential. Anonymity can help to minimize the probability of socially desirable responding. Other strategies, such as the use of computer-administered surveys and randomized response techniques have been suggested in the literature[7]. Recently, the use of web-based surveys has increased for sensitive topics, in which participants interact with a computer to answer questions. In addition, the use of computer-assisted self-interviewing methods has been suggested to minimize socially desirable responding for sensitive questions. However, Kim et al.[15] compared response rates among general social surveys, paper-and-pencil personal interviews, and computer-assisted self-interviewing methods. They found lower response rates for computer-assisted self-interviewing methods compared to other methods.

4. Combining qualitative and quantitative methods (Triangulation)

In a measurement context, combining qualitative and quantitative data within a study can be useful in developing a scale. Triangulation refers to combining multiple methods when studying the same phenomena, to minimize systematic bias in study findings[7]. That is, triangulation may help not only in the development of an instrument, but also with respect to insight about the meaning of concepts. Mitchell[16] proposed four principles when combining qualitative and quantitative data in a study: a) the first is to determine the kind of data needed about the problem and the relevance of the problem to the method chosen should be evident, b) the strengths and weaknesses of each method employed should complement each other, c) methods should be selected on the basis of their relevance to the nature of the phenomena of interest, d) the methodological approach employed should be continually monitored and evaluated to ensure that the first three principles are being followed .

There are several types of triangulation, such as data, investigator, theoretical, methodological, and analysis[17]. Data triangulation may mean collecting data from different groups of subjects in different periods or settings. Investigator triangulation involves multiple investigators for collecting and analyzing data. Theoretical triangulation employs multiple perspectives for the same phenomena. Methodological triangulation can be divided into the within-method and between-method or across-method approach[7]. The within-method approach in the context of measurement may employ several different methods within a single instrument, such as the multitrait-multimethod approach. This within-method approach is useful if the concept has multidimensional properties. The between-method approach refers to the use of both quantitative and qualitative data within the same study. In the measurement context, two or more statistical techniques for analyzing the same set of data can be performed. This is known as data analysis triangulation.

CONCLUSION

Questionnaire design is more of an art than a science. This paper we have tried to help researchers in considerations prior to questionnaire design, steps in development, and relevant details. The importance during the process of questionnaire design is attention to the purpose of the questionnaire. The flow of items should be clear and easy to understand in order to gather precise information. Moreover, when using an existing questionnaire and performing cultural adaptation, psychometric properties and cultural equivalence should be initially evaluated. A pilot test will help to evaluate preliminary questions prior to administration to avoid later mistakes.

REFERENCES

1. Colosi L. Designing an effective questionnaire [Internet]. Ithaca, NY: Cornell University; 2006 [cited 2015 April 27]. Available from: https://www.gateshead.gov.uk/DocumentLibrary/council/consultation/Questionnaire-design-guidance-web.pdf.

2. Frary RB. A brief guide to questionnaire development. Blacksburg, VA: Virginia Polytechnic Institute and State University; 2003.

3. Stehr-Green PA, Stehr-Green JK, Nelson A. Developing a questionnaire. FOCUS on Field Epidemiology. 2003;2(2):1-6.

4. Rattray J, Jones MC. Essential elements of questionnaire design and development. Journal of Clinical Nursing. 2007;16(2):234-243. http://dx.doi.org/10.1111/j.1365-2702.2006.01573.x

5. Netemeyer RG, Bearden WO, Sharma S. Scaling procedures: Issues and applications. Thousand Oaks, CA: Sage Publications; 2003.

6. Leung WC. How to design a questionnaire. Student BMJ: Student British Medical Journal. 2001;9:187-189.

http://dx.doi.org/10.1136/sbmj.0106187

7. Waltz C, Strickland OL, Lenz E. Measurement in nursing and health research. 4th ed. New York, NY: Springer Publishing Company; 2010.

8. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186-3191.

9. Corless IB, Nicholas PK, Nokes KM. Issues in cross-cultural quality-of-life research. Journal of Nursing Scholarship. 2001;33(1):15-20.

10. van de Vijver F, Poortinga YH. Testing in culturally heterogeneous populations: When are cultural loadings undesirable? European Journal of Psychological Assessment. 1992;8(1):17-24.

11. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York, NY: McGraw-Hill; 1994.

12. Knowles ES. Item context effects on personality scales: Measuring changes the measure. Journal of Personality and Social Psychology. 1988;55(2):312-320. http://dx.doi.org/10.1037/0022-3514.55.2.312

13. Gibson D, Wermuth L, Sorensen JL, Menicucci L, Bernal G. Approval need in self-reports of addicts and family members. International Journal of the Addictions. 1987;22(9):895-903.

14. Dijkstra W, Smit JH, Comijs HC. Using social desirability scales in research among the elderly. Quality and Quantity. 2001;35(1):107-115. http://dx.doi.org/10.1023/A:1004816210439

15. Kim J, Kang JH, Kim S, Smith TW, Son J, Berktold J. Comparison between self-administered questionnaire and computer-assisted self-interview for supplemental survey nonresponse. Field Methods. 2009;22(1):57-69. http://dx.doi.org/10.1177/1525822X09349925

16. Mitchell ES. Multiple triangulation: A methodology for nursing science. ANS: Advances in Nursing Science. 1986;8(3):18-26.

17. Kimchi J, Polivka B, Stevenson JS. Triangulation: Operational definitions. Nursing Research. 1991;40(6):364-366.