International Journal of Economics, Commerce and Management
United Kingdom Vol. V, Issue 5, May 2017
http://ijecm.co.uk/ ISSN 2348 0386

COMMON PROCEDURES FOR DEVELOPMENT, VALIDITY AND RELIABILITY OF A QUESTIONNAIRE

Ajay S. Singh
Department of AEM, Faculty of Agriculture, University of Swaziland, Luyengo, Swaziland
singhas64@hotmail.com, asingh@uniswa.sz

Abstract
Questionnaires are the most commonly used data collection methods in applied research for evaluation or assessment of inputs. It is more useful instrument especially in socio-demographic, economic and KAP (Knowledge, Attitude and Practice) studies. This article describes the common steps for development of questionnaire and testing of questionnaires (in terms ofvalidity and reliability) for effective and unbiased outputs. Some common procedures involved in development and testing of questionnaire is research objectives, conceptualization of questionnaire, format and data analysis, and establishing validity and reliability. In this paper also describe the concept of validity and reliability testing for questionnaire. Testing of reliability is based on commonly accepted Cronbach's alpha test. Development of questionnaires is a must to reduce many measurement errors. These common procedures in questionnaire development and testing will enhance data quality and utilization of research outputs. This article also helps to researcher for better understanding of reliability and validity. No doubt, if researchers will consider the validity and reliability of their questionnaire, it could affect the believability of the study and also good impact on the results.

Keywords: Questionnaire, scale, reliability, validity, Cronbach's alpha

INTRODUCTION

Modern applied research based on better living management is quite complex, requiring multiple sets of skills such as agricultural, medical, social, technological, mathematical, statistical etc. Suitable research designs and measurement instrument provide the unbiased estimates of the indicators, conclusions, and their interpretations.

Surveys can be an effective means to collect information or data required for research and evaluation. However, the method is often misused and abused. The challenge is to design a survey that achieves its purpose and avoids the common errors (Measurement Error, Sampling Error, Frame Error, Selection Error, Non-response Error etc.). In this context, questionnaire is primary requirement; it is associated with measurement error.

Questionnaires are the most popular and frequently used technique for data collection in community based applied and agricultural research for evaluation. Questionnaires help gather information on knowledge, attitudes, opinions, behaviours, facts, challenges and other information.

In a review of 748 research studies conducted in agricultural and Extension education, Radhakrishna, Leite, and Baggett (2003) found that 64% used questionnaires. They also reported that a third of the studies reviewed did not report procedures for establishing validity (31%) or reliability (33%). Development of a valid and reliable questionnaire is a must to reduce measurement error. Groves (1987) defines measurement error as the "discrepancy between respondents' attributes and their survey responses".

The importance of measuring the accuracy and consistency of research instruments (especially questionnaires) known as validity and reliability, respectively, have been documented in several studies (Bolarinwa, 2015; Norland-Tilburg,1990;Anastasiadou, 2011; Singh and Masuku, 2012), but their measure is not commonly carried out among health, agriculture and social science researchers in developing countries. This has been linked to the dearth of knowledge of these tests. This comprehensively explores and describes the validity and reliability of a research instrument (with reference to questionnaire).

Development of a valid and reliable questionnaire involves several steps taking considerable attention. This article first describes the steps involved in the development of questionnaires used for data collection. Second, describe the validity and reliability and lastly describe the application of Cronbach's alpha model and interpretation.

BASIC STEPS FOR QUESTIONNAIRE DEVELOPMENT

Background

In this step, the research problems, purpose, objectives, questions, and hypothesis of the proposed research are examined carefully. Determining who is the population or audience, their background, especially their educational levels, access, and the process used to select the respondents are also part of this step. A thorough understanding of the research problem through literature review and readings is very important. Good homework and understanding of

Step1 provides the strong foundation for next Step.

What is questionnaire(Gault,1907)?

A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. Although they are often designed for statistical analysis of the responses, this is not always the case. The questionnaire was invented by the Statistical Society of London in 1838.

Usually, a questionnaire consists of a number of questions that the respondent has to answer in a set format. A distinction is made between open-ended and closed-ended questions. An open-ended question asks the respondent to formulate his own answer, whereas a closed-ended question has the respondent pick an answer from a given number of options. The response options for a closed-ended question should be exhaustive and mutually exclusive.

Conceptualization of Questionnaire

After developing a good understanding of the research problems and objectives, the next step is to make or generate the questions or statements for the questionnaire. In this step, content is transformed into statements/questions on the basis of theoretical framework and review literature. In addition, a link among the objectives of the study and their translation into content is established. For example, the researcher must highlight what the questionnaire is measuring, that is, knowledge, attitudes, perceptions, opinions, practice, recalling facts, behaviour change, etc. Major variables (dependent and independent variables) are identified and explained in this step.

Research Questions: Questions must be clear and associated with research objectives in simple language.

Decide what you are measuring: As with determining the purpose, this should be based on the objectives of your research and the evaluation of its results and impact.

Arrange in a logical order: Group similar questions together, use an outline approach and use number for each question.

Minimize open-ended questions: Open ended questions are difficult to score and summarize. Field testing a questionnaire with open-ended questions can help identify common answer categories that could be made into fixed-response (multiple choices) questions.

Provide space to respondent: Give respondent space to comment about individual questions.

Avoid biased questions: Minimize “bias” in questions by using experts and field testing.

Ask only one question at a time: Avoid double minded questions that confuse the respondent into not knowing how to answer.

Looks smart and professional: Questionnaire should be smart and objective oriented without any typing mistakes. Keep the questionnaire as short as possible. Focus on questions and minimize the required time for information, include definitions if needed. Use simplest language necessary and avoid jargon and acronyms

Enclosed a cover letter: A good covering letter enclosed with questionnaire, including all necessary information with purpose of study and its usefulness. Why response is important and promise of confidentiality and explanation of identification and also mention deadline for the survey.

Thank respondents: Lastly thanks to respondent on questionnaire and in covering letter.

Format and Data Capturing and Analytical Approach

In this step, the focus is on writing statements and questions, selection of appropriate measurement scales, questionnaire layout, format, order of questions, very careful about the qualitative and quantitative variable, coding, font style and size (should be attractive), front and back cover, and proposed data analysis. Scales are devices used to quantify a subject's response on a particular variable. Understanding the relationship between the level of measurement and the appropriateness of data capturing and analysis is very important. For example, analysis of variance is one mode of data analysis, the independent variable must be measured on a nominal scale with two or more levels (yes, no, not sure/ applicable, not applicable etc.), and the dependent variable must be measured on an interval/ratio scale (strongly agree, agree, disagree , strongly disagree).

Validity and Reliability of Questionnaire

Validity and reliability are concepts of measurement criteria of a survey, questionnaire or any another type of measure. Validity and reliability are important concepts in research. The everyday use of these terms provides a sense of what they for example; opinion is valid or not and your instrument is reliable or not. In research, however, their usein research is more complex.

To assess the validity and reliability of a survey or other measure, researchers need to consider a number of things. If researchers are considering these points of validity and reliability of the questionnaire, it could affect the believability of the study.

Researchers and Scientists are continuously working on validity and reliability issues. Carmines and Zeller (1979) discussed about the reliability and validity. Winer et al. (1991) and Dunn et al. (2013), described the criteria and practical solution to the pervasive problem of validity, reliability and internal consistency estimation.

What is Validity?

The validity of a study is dependent on the degree of systematic error. Validity is usually separated into two components (Singh and Masuku, 2012).

There are two types of validity, known as internal and external validity. Internal validity is dependent on the amount of error in measurements, including exposure of disease and the associations between the variables. Good internal validity implies a lack of error in measurement and suggests that inferences may be drawn. External validity pertains to the process of generalizing the findings of the study to the population from which the sample was drawn. It requires an understanding of which conditions are relevant to the generalization (Singh and Masuku, 2012).

Validity of Questionnaire:

Face validity: Researchers need to consider theface validity of a questionnaire. That is, to a layperson, does it look like it will measure what it is intended to measure? In our example, would the people administering and taking the questionnaire think it a valid measure? Do the questions and range of response options seem, on their face, appropriate for measuring objectives?

Content validity: Researchers also need to consider the content validity of the questionnaire; that is, will it actually measure what it is intended to measure. Researchers frequently trust on subject-matter experts to help determine this.

Construct validity: This refers to the questionnaire's ability to measure the concept adequately. When questionnaires are measuring something conceptual or theoretical, researchers also important to establish its construct validity. In this case, the researchers could have given a questionnaire on a similar concept, to see if the results were related, as one would expect. Or researcher could have given a questionnaire on a diverse concept, to see if the results were the opposite.

Criterion validity or Predictive Validity: Criterion validity; that is, the extent to which the measurement tool is able to produce accurate findings when compared to a standardized finding. It may sometimes be appropriate for researchers to establish criterion validity. In other words, Expectation of future performance based on the results obtained currently by the measure; correlate the scores obtained with the performance. The later performance is called the criterion and the current score is the prediction. This is an empirical check on the value of the test – a predictive validation.

Concurrent Validity: Concurrent validity is the degree to which the results or scores on a test are related to the scores on another, already established, test administered at the same time, or to

some other valid criterion available at the same time. Logically, predictive and concurrent validation are the same, the term concurrent validation is used to indicate that no time forgotten between measures.

Construct Validity: Construct validity is the degree to which a test measures an intended hypothetical construct. Several times researchers evaluate the abstract attributes (qualitative characteristics) or constructs. The process of validating the interpretations about that construct as indicated by the test score or results is known as construct validation.

Reliability of questionnaire:

Test-retest reliability: Researchers also need to consider the reliability of a questionnaire. If the researchers get similar results if they repeat their questionnaire soon after and conditions have not changed? This is called test-retest reliability.

Internal consistency: Internal consistency is another aspect of reliability concerns with consistency among the questions. Do similar questions give rise to similar answers? For example, if two questions are related to effect of fertilizer, the researchers would expect the responses to be consistent.

Inter-rater reliability: Researchers also consider the inter-rater reliability; that is, would different individuals assessing the same thing score the questionnaire the same way.

Equivalent-Forms or Alternate-Forms Reliability: Two tests that are similar in every manner except for the actual items included in the questionnaire. Used when it is likely that test takers will recall responses made during the first session and when alternate forms are available. Correlate the two scores. The obtained coefficient is called the coefficient of stability or coefficient of equivalence.

Split-Half Reliability: In this case requires only one administration, especially appropriate when the test is very long. The most commonly used method to split the test into two is using the odd-even strategy. Since longer tests tend to be more reliable, and since split-half reliability represents the reliability of a test only half as long as the actual test, a correction formula must be applied to the coefficient. Split-half reliability is a form of internal consistency reliability.

Rationale Equivalence Reliability: Rationale equivalence reliability is not established through correlation but rather estimates internal consistency by determining how all items on a test relate to all other items and to the total test.

Internal Consistency Reliability: Determining how all items on the test relate to all other items. Kudser-Richardson formula is an estimate of reliability that is essentially equivalent to the average of the split-half reliabilities computed for all possible halves.

Kudser-Richardson KR-20 (Allen and Yen,1979):

KR-20 = n/(n − 1)[1 − Sum(Pi Qi)/Var(X)]

Where:
n = Total number of items
Sum (Pi Qi) = Sum of the product of the probability of alternative responses
Qi = 1- Pi
Var(X) = Composite variance.

And to calculate coefficient alpha (a):

α= n/(n − 1)[1 − Sum Var (Yi)/Var (X)].

Where:
n = Number of items
Sum Var(Yi) = Sum of item variances
Var(X) = Composite variance.

Standard Error of Measurement: Reliability can also be expressed in terms of the standard error of measurement. It is an estimate of how often you can expect errors of a given size.

Reliability Test:
How to test reliability of questionnaire and interpret the test result is also challenging for researchers. Some important techniques used by the researcher according the convenience.
* Spearman-Brown prediction formula: This test is useful for Split-half reliability test. In some situations may require extremely precise questionnaire with very high reliabilities. In the extreme case of a two-item test, the Spearman–Brown prediction formula is more appropriate (Eisinga, et al. 2013).
* Kudser-Richardson test: This is useful for internal consistency reliability of items in the questionnaire. It is also known as Kuder–Richardson Formula 20 (KR-20), which is an equivalent measure for dichotomous items.
* Cronbach's alpha test: This important and more useful test for internal reliability of questionnaire. It is one way concept of measuring strength of that consistency.
What is Cronbach's alpha?
Cronbach's alpha statistic is widely used in the social sciences, business, nursing, and other disciplines.

Cronbach's alpha is a measure used to assess the reliability, or internal consistency, of a set of scale or test items. In other words, the reliability of any given questionnaire refers to the extent to which it is a consistent measure of a concept, and Cronbach's alpha is one way of measuring the strength of that consistency(Cronbach,1951).

Cronbach's alpha is computed by correlating the score for each scale item with the total score for each observation (usually individual survey respondents), and then comparing that to the variance for all individual item scores.

Cronbach's alpha test Formula (Cronbach,1970):

Suppose that X = Y1 + Y2 + Y3 + …… Yk

Cronabach's α is defined as

$$\alpha=(K/K–1)(1–\sum\{Var(Yi)/Var(X)\};i= 1 \text{ to } k$$

Where,

Kis the number of scale items, V(Yi) refers to the variance associated with item i and V(X) refers to the variance associated with the observed total scores.

If the items are scored 0 and 1

Cronabach's α is defined as

$$\alpha = (K/K–1)\ [\ 1- \{\sum Pi\ Qi / V(X)\}] ;i= 1 \text{ to } k$$

Cronbach's alpha can also be defined as:

$$\alpha=k \times \mu / \{\lambda+ (k–1)\ \mu\}$$

Where,

K refers to the number of scale items, μ refers to the average of all covariance between items and λ refers to the average variance of each item.

Therefore, Cronbach's alpha is a function of the number of items in a test, the average covariance between pairs of items, and the variance of the total score.

Standardized α will be defined as

$$\alpha \text{ stand.} = K\ \delta / \{\ 1+ (K – 1)\ \delta\ \}$$

Where,

δ is the mean of K (K – 1)/2 non redundant correlation coefficient.

The theoretical value of alpha varies from zero to one, since it is the ratio of two variances. However, depending on the estimation procedure used, estimates of alpha can take on any value less than or equal to 1, including negative values, although only positive values make

sense (Eisinga, et al. 2013).Higher values of alpha are more desirable. Some professionals, as a rule of thumb, require a reliability of 0.70 or higher (obtained on a substantial sample) before they will use an instrument. Although Nunnally (1978) is often cited when it comes to this rule, he has actually never stated that 0.7 is a reasonable threshold in advanced research projects. And obviously, this rule should be applied with caution when α has been calculated from items that systematically violate its assumptions, such as the use of ordinal items, as Cronbach's alpha is only a lower limit for a large sample within a metric space.

Cronbach's alpha	Internal consistency
α ≥ 0.9	Excellent
0.8> α ≥ 0.9	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 > α	Unacceptable

The value of alpha and a sample with a narrow range can deflate it, so this rule should be used with caution (Cortina,1993).

A commonly accepted rule for describing internal consistency using Cronbach's alpha is as follows, [Green et al. 1977; Kline, 2000; DeVellis, 2012)though a greater number of items in the test can artificially inflate the value of alpha and a sample with a narrow range can deflate it, so this rule should be used with caution (Nunnally,1978).

Interpretation of Cronbach’s alpha:

The resulting α coefficient of reliability ranges from 0 to 1 in providing the overall assessment of a measure’s reliability. If all of the scale items are independent from one to another (i.e., are not correlated), then α = 0; and, if all of the items have high correlated, then α will approach to 1 as the number of items in the scale tends to infinity. In other words, the higher the α coefficient, the items have common covariance and probably measure the same concept.

Although the standards for what makes a good α coefficient are entirely arbitrary and depend on researcher theoretical concept of the scale in question, many researchers recommend a minimum α coefficient between 0.65 and 0.8 (or higher) and α coefficients that are less than 0.5 are generally not acceptable.

In interpretation a scale’s α coefficient, remember that a high α is both a function of the covariance among items and the number of items in the analysis, so a high α coefficient isn’t in and of itself the mark of a good or reliable set of items; researcher can often increase the α

coefficient by increasing the number of items in the analysis and highly correlated items will also produce a high α coefficient, if it's very high (> 0.95), may be risking idleness in scale items.

If all of the scale items you want to analyse are binary and you compute Cronbach's alpha, you're actually using an analysis called the Kuder-Richardson KR-20. The formula for Cronbach's alpha builds on the KR-20 formula to make it suitable for items with scaled responses (Likert scaled items) and continuous variables, so the underlying math is, if anything, simpler for items with dichotomous response options. After using this test, you'll get the same α coefficient and other similar output, and you can interpret this output in the same ways described above.

Limitations of Cronbach's Alpha reliability test(Ritter, 2010; Green et al., 1977):
Cronbach's alpha is not a measure of dimensionality, nor a test of unidimensionality. In fact, it's possible to produce a high α coefficient for scales of similar length and variance, even if there are multiple underlying dimensions. To check for dimensionality, you'll perhaps want to conduct an exploratory factor analysis. Cronbach's Alpha reliability test is not robust against missing data or observations.

Cronbach's alpha is also not a measure of validity, or the extent to which a scale records the "true" value or score of the concept and trying to measure without capturing any unintended characteristics. For example, word problems in an algebra class may indeed capture a student's math ability, but they may also capture verbal abilities or even test. A reliable measure is one that contains zero or very little random measurement error—i.e., anything that might introduce arbitrary or haphazard distortion into the measurement process, resulting in inconsistent measurements. However, it need not be free of systematic error—anything that might introduce consistent and chronic distortion in measuring the underlying concept of interest in order to be reliable; it only needs to be consistent.

Example based on reliability test (Cronbach's Alpha reliability test):
Anastasiadou (2011) developed the questionnaire / instrument, which intended to measure students' attitudes towards statistics, is Students Attitudes toward Statistics and Technology Scale (SASTSc). This tool consisted of 28 items referring to five different attitude subscales, as follows: (a) Statistics Cognitive Competence-positive and negative attitudes concerning a student's knowledge and skills as applied to statistics (Co1, Co2, Co3, Co4, Co5, Co6); (b) Technology Cognitive Competence-positive and negative attitudes concerning a student's knowledge and skills as applied to technology –computers (Te1, Te2, Te3, Te4); (c) Attitudes to learning statistics with technology (ST1, ST2, ST3, ST4, ST5, ST6) -positive and negative

attitudes concerning a student's attitudes to learning statistics with technology; (d) Value- positive and negative attitudes to the worth and usefulness of statistics in students' personal and professional life (Va1, Va2, Va3, Va4, Va5, Va6); (e) Affect- positive and negative emotions concerning statistics (Af1, Af2, Af3, Af4, Af5, Af6). The 28 items have created the above 5 different attitude subscales, thus those subscales are the results of the explanatory factor analysis. Each item of the instrument used a 5-point Likert scale that ranged from 1- Strongly Disagree to 5-Strongly Agree. The value of the Cronbach's α coefficient for this instrument in this study's sample was 0.901.

CONCLUSIONS

Research investigation is the part of an extensive development with regard to community health, finance, education, and agriculture, etc. that are indicators of better life of human beings. Any social phenomenon and especially those that can be characterized by numerical facts are the results of one or more causes of interaction. In this context, questionnaire and it is one of the most important aspects for good results. Validity and reliability condition is essential part for questionnaire.

Therefore, development of the questionnaire for data collection is important to reduce measurement errors-questionnaire content, questionnaire design and format, and respondent. Well-designed conceptualization of the content and transformation of the content into questions is essential to minimize the error. Careful attention to detail and understanding of the process involved in developing a questionnaire are of immense value to extension educators, graduate students, and faculty alike. Not following appropriate and systematic procedures in questionnaire development, testing and evaluation may undermine the quality and utilization of data (Esposito, 2002). Anyone involved in research and evaluation research, must, at a minimum, follow these important points to develop a valid and reliable questionnaire to improve the quality of research.

This paper provided the basic knowledge about development of questionnaire, validity and testing of reliability. More importantly, we have discussed these measures and provided how we should use these measures in our day-to-day research investigation and also have illustrated how to calculate reliability of questionnaire. A careful consideration of development, reliability and validity of questionnaire will hopefully result in more meaningful studies whose results and interpretations are based on sound scientific principles.

REFERENCES

Allen M. J. and Yen, W. M. (1979).Introduction to Measurement Theory. Monterey, CA: Brooks/Cole.

Anastasiadou, S. D. (2011). Reliability and validity testing of a new scale for measuring attitudes towards learning statistics with technology, ActaDidacticaNapocensia, 4(1).

Bolarinwa O.A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Nigerian Postgraduate Medical Journal, 22,195-201.

Carmines, E., and Zeller, R. (1979).Reliability and Validity Assessment.Sage Publications, Beverly Hills, California.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications.Journal of Applied Psychology.78: 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.Psychometrika.16 (3): 297–334.

Cronbach, L. J. (1970). Essential Psychological Testing, Harper & Row, p. 161.

DeVellis, R.F. (2012). Scale development: Theory and applications. Los Angeles: Sage. pp. 109–110.

Dunn, T. J., Baguley, T. and Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology.

Eisinga, R., TeGrotenhuis, M. and Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach or Spearman-Brown?.International Journal of Public Health.58 (4): 637–642.

Esposito, J. L. (2002). Interactive, multiple-method questionnaire evaluation research: A case study. Paper presented at the International Conference in Questionnaire Development, Evaluation, and Testing (QDET) Methods. Charleston, SC.

Gault, R. H. (1907). A history of the questionnaire method of research in psychology.Research in Psychology.14 (3): 366–383.

Green S. B.,Lissitz, R. W. and Mulaik S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement.37: 827–838.

Groves, R. M. (1987). Research on survey data quality.Public Opinion Quarterly, 51, 156-172.

Kline, P. (2000). The handbook of psychological testing (2nd ed.). London: Routledge, page 13

Norland-Tilburg, E. V. (1990). Controlling error in evaluation instruments.Journal of Extension, 28(2).

Nunnally, J. C. (1978). Assessment of Reliability. In: Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Radhakrishna, R. B. Francisco, C. L. and Baggett. C. D. (2003). An analysis of research designs used in agricultural and extension education.Proceedings of the 30th National Agricultural Education Research Conference, 528-541.

Ritter, N. (2010). Understanding a widely misunderstood statistic: Cronbach's alpha. Paper presented at Southwestern Educational Research Association (SERA) Conference 2010: New Orleans, LA (ED526237).

Singh, A.S. and Masuku, M. B. (2012). Understanding and applications of test characteristics and basic inferential statistics in hypothesis testing, European Jr. of Applied Sciences, 4(2), 90-97

Winer, B., Brown, D., and Michels, K.(1991).Statistical Principles in Experimental Design, Third Edition. McGraw-Hill, New York.