

On Survey Questionnaire Testing

Rebeka Lea Sajko

Faculty of Medicine, University of Ljubljana, Slovenia

Correspondence: *rebeka.sajko@gmail.com*; + 386 2 4622777

Received: July 28 2024; **Accepted:** September 9 2024

Abstract

The purpose of survey questionnaire testing is to ensure that the questionnaire is effective, reliable, and valid research instrument. It should accurately reflect the intended construct and provide stable and consistent results. In recent years, there has been a growing emphasis on the importance of survey questionnaire testing; however, relatively little methodological research and guidance about testing methods is provided. Our goal is to present the methods available for questionnaire testing in brief and give some advice on how to use them. **Conclusion** – Testing a survey questionnaire before its deployment is crucial in ensuring the data's accuracy and consistency. There are various testing methods available that address the content and technical quality of the research instrument.

Key Words: Validity ■ Reliability ■ Research Instrument.

Introduction

Surveys are frequently used to gather data in the fields of health and social sciences. They can be carried out through online forms, telephone interviews, face-to-face discussions, or mailed questionnaires. Regardless of the method of administration, it is essential to have well-crafted questions that are clear to the intended audience and that focus on the concepts the researcher wants to explore. Social science education, such as in the curriculum for social informatics students, includes some instruction on survey methodology. However, researchers in the natural sciences typically encounter survey methodology only later in their careers, when they are already faced with survey research and are expected to formulate research objectives and design survey questions properly. Often no pre-existing survey instruments are available, so researchers have to develop a questionnaire on their own. This is where challenges can arise.

A well-structured and consistent questionnaire is essential for collecting relevant data in line with

the research concept. Hence, it is crucial to test the questionnaire for its suitability before deploying it. Questionnaire testing dates back to the late 1930s, with the first publications on the subject appearing in early 1940s. However, until 2004, there was no monograph dedicated exclusively to survey questionnaire testing and evaluation (1).

In recent years, there has been an increasing emphasis on the importance of testing the survey questionnaire to ensure high-quality collected data. However, there are still no clear guidelines on how to test the questionnaire before it is fielded to ensure that all survey questions meet three basic standards: 1) content standards (e.g., are the questions asking about the right things?); 2) cognitive standards (e.g., do respondents understand the questions consistently; do they have the information required to answer them; are they willing and able to formulate answers to the questions?); and 3) usability standards (e.g., can respondents and interviewers complete the questionnaire easily and as they were intended to?) (2).

In our article, we briefly reviewed the methods available for testing the validity and reliability of questionnaires and offered some advice on how to use them.

Validity and Reliability Testing

Two important qualities of a questionnaire to be tested are accuracy and consistency. These are assessed by considering the questionnaire's validity and reliability. There are various definitions of validity and reliability within survey research. To list a few common definitions: "the extent to which the survey measure accurately reflects the intended construct" (2), or the extent to which an instrument (e.g., a questionnaire) measures what it is supposed to measure, or an assessment of the instrument's accuracy, etc.

Reliability, on the other hand, is defined as "a measurement of variability of answers over repeated conceptual trials" and it "refers to the consistency of measurement either across occasions or across items designed to measure the same construct" (2), or simply, it measures the consistency of research instrument. In the literature, you can find various classifications of methods for testing survey questionnaires, and it's important to note that the terminology is not standardized. As such, some terms related to testing methods are often used interchangeably.

One way to classify questionnaire testing is by distinguishing between two types: expert and practical testing. In expert testing (or expert evaluation), specialists assess the survey questionnaire from a content perspective (content experts) and a technical perspective (survey methodology experts). Ideally, both types of experts should be involved in the evaluation, or at least, an expert in one area should have some understanding of the other (personal communication with a professional statistician). After completing the content testing, the researcher makes appropriate corrections to the questionnaire at their discretion. The next step is practical evaluation, which involves pilot testing the survey on a small number of individuals who ideally represent the population on which the survey

will be conducted. Based on the results of the practical evaluation, the researcher makes adjustments to the questionnaire as they see fit. With this, the testing phases are completed, and the questionnaire is ready to be fielded.

Questionnaire testing can also be divided into qualitative and quantitative methods (Table 1). Various testing options are presented, although typically not all are used for testing an individual questionnaire. One method or a combination of several methods can be employed. The decision lies with the researcher and generally depends on the purpose of the survey, the complexity of the questionnaire, and ultimately, the researcher's personal preferences.

To briefly explain the testing methods shown in Table 1:

Preliminary pilot testing: as a part of the questionnaire developmental phase (also shown in Fig. 2). Before conducting a pilot test of the questionnaire on the intended respondent group, it is advisable to test the questionnaire on a small sample of respondents to gather their suggestions for potential improvements to the items. The questionnaire items should be revised upon reviewing the results of the preliminary pilot testing (3).

Focus group: Focus groups are usually small groups of subjects who share similar characteristics with the target population. This method can be applied before the developmental phase or after the development of a questionnaire, i.e., during the testing phase. Focus groups are efficient ways of gathering qualitative information about the survey topic from the perspective of the target population prior to imposing the structure of a survey questionnaire (2).

Face validity: Face and content validity are subjective opinions of non-experts and experts. Face validation can be determined using dichotomous scale yes-no and assessed by the Cohen's Kappa Index.

Cognitive testing (also known as cognitive interviews): It is assessed whether the questions are understandable. This is particularly suitable for preparing surveys intended for the general population, where it is expected that different respondents

will understand the questions to varying degrees. Typically, cognitive testing is conducted in the form of interviews with either a group of individuals or a group of experts.

Content validity: Content validation is done either by a group of non-experts or by a panel of content experts. It refers to the extent to which the items in a questionnaire are representative of the theoretical construct.

Construct validity: A parameter designed to measure the theoretical construct that it is intended to evaluate. Construct validity involves testing a questionnaire by estimating the association of that construct with other variables. Construct validity can be expressed using a correlation coefficient. It can be tested also by factor analysis and principal component analysis to identify the underlying components.

Expert evaluation (also known as Delphi method): This involves obtaining expert consensus on the questions: the questionnaire is sent to a group of experts in several sequential phases. The experts review the questionnaire, provide their feedback, and the questionnaire is then revised based on this feedback and sent back to the experts for further evaluation, and so on. There are two types of

expert evaluations: a) content evaluation (by content experts), and b) technical evaluation (by survey methodology experts). The questionnaire is typically sent for evaluation to one of these groups of experts (though occasionally to both, especially for large and complex studies). A sufficient number of experts can range from as few as one or two to about five, depending on the complexity of the study.

Pilot testing: Pretesting on a subset of the intended population to gather suggestions for potential improvements to the questions. The questionnaire items should be revised upon reviewing the results of the pilot testing. The optimal sample size for pilot testing is not clearly defined. Ideally, the sample should be as representative as possible of the target population. The size of the pilot test sample generally depends on the complexity of the survey – smaller surveys might only require 20-30 participants, with some recommendations suggesting up to 50 participants. For more complex surveys, it is often recommended to have a minimum of 200 participants, although some guidelines suggest as much as 10% of the target population, which could be a very large number of subjects.

Table 1. Overview of Questionnaire Testing Methods

Testing method	Questionnaire characteristic to be measured			Type of testing		Quality measure		Testing group	
	Q1	Q2	Q3	Qualitative	Quantitative	Validity	Reliability	Expert	Respondent
Preliminary pilot testing	✓	✓	-	✓	-	✓	-	-	✓
Focus group	✓	✓	-	✓	-	✓	-	-	✓
Face validity	✓	✓	-	✓	-	✓	-	✓	✓
Cognitive testing	✓	-	-	✓	-	✓	-	✓	✓
Content validity	-	✓	✓	✓	-	✓	-	✓	✓
Construct validity	-	-	✓	✓	✓	✓	-	✓	-
Expert evaluation	-	✓	-	✓	-	✓	-	✓	-
Pilot testing	✓	✓	✓	✓	✓	✓	-	-	✓
Internal consistency	-	-	-	-	✓	-	✓	-	✓
Test-retest reliability	-	-	-	-	✓	-	✓	-	✓
Inter-rater reliability	-	-	-	-	✓	-	✓	✓	✓

Q1: How well the questions are understood? How difficult are they to answer?; Q2: Are the questions asking about the right things? Do they provide answers to our objectives?; Q3: Are the questions constructed properly? Are the items representative of the entire theoretical construct?.

The researcher's decision on sample size will mainly depend on the time available for conducting the survey, the complexity of the study, and the budget. If multiple testing methods are used simultaneously, a very large sample for pilot testing may not be as critical.

Internal consistency: Defined as the consistency in the measurement of the intended construct. It is a measure of the inter-correlation of the questionnaire items, commonly calculated by Cronbach's alpha coefficient.

Test-retest reliability: In test-retest reliability assessment, the questionnaire is administered to the same group of respondents at different time points and repeating the research. The measure is evaluated using either a correlation coefficient or intra-class correlation coefficient (ICC).

Inter-rater reliability (also known as inter-rater agreement): Measures the agreement between subjective ratings by multiple raters, either experts or non-experts. This consistency can be estimated using Kappa statistic.

In Fig. 1., the process of conducting survey research is illustrated, emphasizing the crucial phases involved in questionnaire development and implementation. Testing the questionnaire is just an aspect of the comprehensive process. Prior to designing and testing survey questions, the *initial or planning phase* must be completed, which involves clearly defining the study's focus and formulating the concepts that need to be measured.

Once objectives have been clearly stated, questions are designed in the *developmental phase*. A preliminary pilot test (as described above) can be conducted, but it is not mandatory. This test provides the researcher with the first insight into respondents' attitudes towards the questions. If necessary, the questions in the first draft of the questionnaire are corrected, eliminated, or added. The *testing phase* can be part of the developmental phase or a subsequent phase. It includes all testing methods as illustrated in Table 1.

However, typically, only some of these methods are used for the validation of a specific questionnaire, as illustrated in Fig. 1.

After finalizing, the survey questionnaire is *implemented* and administered to the intended group of respondents. The collected data is *analyzed* and interpreted according to the study objectives. The final step in survey research is *reporting* the results. It is advisable to prepare the study report in line with established recommendations, such as CROSS guidelines (4) listed on the EQUATOR Network.

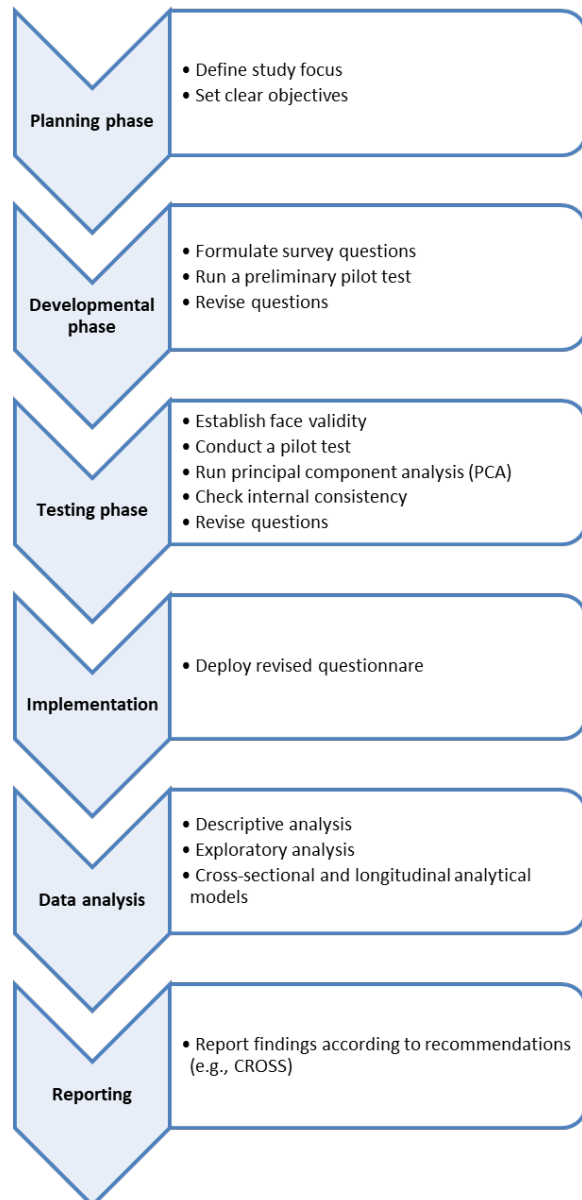


Fig. 1. Steps in the survey research.

Discussion

Testing a questionnaire is not necessarily required if it contains questions that have already been tested in previous studies. However, opinions on this matter are divided within the professional community. Some experts in survey methodology argue that this approach is preferred, as developing a new questionnaire is a complex task, and it should be regarded as a measurement tool (similar to scales, thermometers, etc.) used to gather subjective responses such as opinions and attitudes. On the other hand, some experts believe that, since each questionnaire has unique characteristics that are ultimately validated in a specific population, certain aspects of it should be re-examined in each study. Nevertheless, it is often the case that there is no suitable existing questionnaire for the research, and we must develop a questionnaire on our own. Before finalizing and implementing the questionnaire, it should be tested for suitability. We should assess whether the questions are understandable and whether they measure our construct and provide answers to our objectives.

The choice of testing method depends on the researcher's judgment. They may decide to test the questionnaire using only one or multiple methods. Often, researchers opt only for expert review without conducting a pilot survey (personal communication with an expert in survey methodology). This is particularly the case when the survey aims to identify potential background factors / latent structures, and pre-structuring the questions into sets would not be sensible. However, when the research concept is designed to explore respondents' opinions on pre-established thematic sets of questions, it makes sense first to determine the consistency of the individual sets of questions and adjust them as appropriate before finalizing the questionnaire and offering it to the intended respondent group.

When preparing an article based on survey methodology, it is highly recommended to use one of the established guidelines for reporting on cross-sectional studies (4–6). However, there are currently very few published articles based on

survey research design that cite such guidelines or have been prepared in accordance with them. The guidelines CHERRIES, published in 2004 (5), are designed for web-based surveys. In the checklist of these guidelines, the testing is mentioned in one sentence, as testing “the usability and technical functionality of the electronic questionnaire.” Furthermore, the SURGE guidelines from 2014 (6), are intended for self-administered postal surveys. The checklist of the guidelines includes only a question on whether a pre-test was provided and whether the reliability and validity of the tool have been reported. Analogous to several important guidelines for various types of research studies on the EQUATOR Network website (for example, the CONSORT statement for randomized trials), a universal guideline called CROSS was created in 2021 and is listed on this website as a standard quality-reporting tool for surveys (4). Information on pretesting is included in the checklist under the topic 'Data collection methods' and includes the following recommendations: a) provide information on pretesting of the questionnaire, if performed; b) report the method of pretesting, the number of times the questionnaire was pre-tested, the number and demographics of participants used for pretesting, and the level of similarity of demographics between pre-testing participants and the sample population. Recommendations for further reading on survey methodology and on methods for testing and evaluating survey questionnaires: (1, 2).

Conclusion

In recent years, there has been a growing emphasis on the importance of survey questionnaire testing. However, clear guidelines on which testing methods to use for a specific survey study design are still lacking. The decision on the method for testing the suitability of a questionnaire rests with the researcher. Recognizing the importance of questionnaire testing in the development and implementation process is crucial to ensure high-quality data that can yield meaningful answers to the research questions.

Conflict of Interest: The author declares that she has no conflict of interest.

Acknowledgement: Hereby, the author would like to express her gratitude to Dr. Ana Slavec, expert in survey methodology and researcher at the InnoRenew CoE in Izola, Slovenia, for her valuable advice.

References

1. Presser S, Couper MP, Lessler JT, Martin E, Martin J, Rothgeb JM, et al. *Methods for Testing and Evaluating Survey Questions*. New Jersey: John Wiley & Sons, Inc.; 2004.
 2. Groves R, James M, Lepkowski, Roger Tourangeau, Eleanor Singer, Floyd J, Fowler Jr., Mick P, Coupe. *Survey methodology*. 2nd ed. New Jersey: John Wiley & Sons, Inc.; 2009.
 3. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth*. 2017 May;11(Suppl 1):S80-9.
 4. Sharma A, Minh Duc NT, Luu Lam Thang T, Nam NH, Ng SJ, Abbas KS, et al. A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). *J Gen Intern Med*. 2021 Oct;36(10):3179-87.
 5. Eysenbach G. Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res*. 2004 Sep 29;6(3):e34.
 6. Grimshaw J. SURGE (The SURvey Reporting GuidelinE). In 2014. p. 206-13.
-