



# Questionnaire Validation and Reliability Testing in Health and Public Health Research

Dr. Shabna G S<sup>1</sup>

<sup>1</sup>Senior Resident, Department of Public Health Dentistry, Government Dental College, Thrissur, Kerala, India  
Corresponding Author: <sup>1</sup>Shabna G S, Senior Resident, Department of Public Health Dentistry, Government Dental College, Thrissur, Kerala, India

Date of Submission: 03-05-2026

Date of Acceptance: 13-05-2026

**ABSTRACT:** Questionnaires are widely used research instruments in health sciences, social sciences, psychology, education, and public health because they provide a practical, economical, and efficient method for collecting data from large populations. However, the accuracy and credibility of findings obtained through questionnaires depend on the psychometric quality of the instrument. Therefore, proper assessment of validity and reliability is essential before the application of a questionnaire in research or clinical practice. Validity refers to the extent to which a questionnaire measures the intended construct, whereas reliability reflects the consistency, stability, and reproducibility of the measurements obtained. This review article discusses the fundamental concepts, types, methods, and statistical approaches used in evaluating questionnaire validity and reliability. Different forms of validity, including face validity, content validity, criterion validity, and construct validity, are described along with commonly used indices such as the Content Validity Index (CVI). Similarly, methods of reliability assessment including test-retest reliability, inter-rater reliability, and internal consistency using Cronbach's alpha are explained.

**KEYWORDS:** Development, Questionnaires, Translation, Validation

## I. INTRODUCTION

Questionnaires are among the most extensively used instruments for data collection in health sciences, social sciences, psychology, education, and public health research. They are frequently employed either as independent tools or in combination with qualitative and observational methods in mixed-method studies. Questionnaires facilitate the systematic collection of information related to knowledge, attitudes, beliefs, perceptions, behaviors, and practices from large populations in a cost-effective and time-efficient manner. Their ability to standardize responses and permit comparisons across different groups and settings has made them an indispensable component of epidemiological and clinical research [1].

Despite their widespread use, the scientific quality of findings generated through questionnaires depends largely on the psychometric strength of the instrument. A poorly designed questionnaire can produce inaccurate, inconsistent, and biased data, thereby compromising the validity of research outcomes. Consequently, careful development, validation, and reliability assessment of questionnaires are essential before their implementation in research and clinical settings. The process of questionnaire validation ensures that the instrument accurately measures the intended construct, while reliability assessment determines the consistency and reproducibility of responses obtained from the tool [2].

Traditionally, questionnaire-based research has been influenced by a positivist approach in which questionnaires are viewed as standardized instruments designed to objectively measure variables and generate quantifiable data. However, contemporary perspectives recognize questionnaires as interactive processes between researchers and participants, where responses may be shaped by social, cultural, psychological, and contextual factors. As a result, questionnaire responses may not always reflect actual behavior but rather self-reported perceptions and socially desirable responses. This limitation is particularly relevant in studies involving sensitive topics such as substance use, sexual practices, dietary behavior, and health-related habits [3].

Validity and reliability are considered the two most important psychometric properties of a questionnaire. Validity refers to the extent to which an instrument measures what it is intended to measure, whereas reliability refers to the degree of consistency and stability of measurements over time and across different conditions. Different forms of validity, including content validity, face validity, criterion validity, and construct validity, are used to assess the appropriateness and accuracy of questionnaire items. Similarly, reliability is commonly evaluated using methods such as test-retest reliability, inter-rater reliability, split-half



reliability, and internal consistency measures including Cronbach's alpha [3,4].

Considering the growing use of questionnaires in research and the importance of psychometric evaluation, a comprehensive understanding of validity and reliability is essential for researchers.

## II. VALIDITY AND RELIABILITY OF THE RESEARCH QUESTIONNAIRE

Validity and reliability are fundamental psychometric properties that determine the scientific quality and credibility of a research instrument. Validity refers to the extent to which a questionnaire or measurement tool accurately assesses the concept or construct it is intended to measure. In addition to measuring the intended attribute correctly, validity also reflects how well the findings obtained from the study participants represent the actual characteristics of the target population. A valid instrument therefore produces results that are meaningful, accurate, and appropriate for the purpose of the study [5].

Reliability, on the other hand, refers to the consistency, stability, and reproducibility of measurements obtained using a research instrument. A reliable questionnaire yields similar results when administered repeatedly to the same individuals under comparable conditions. Variations in responses obtained during repeated assessments or differences between observers administering the same instrument may indicate poor reliability. Thus, reliability reflects the precision and dependability of the measurement process [6].

Although validity and reliability are closely related concepts, they are not synonymous and may exist independently. An instrument may demonstrate high reliability without being valid. For example, a defective weighing scale that consistently records a person's weight as 10 kilograms higher than the actual value provides measurements that are reproducible and consistent; however, the measurements are inaccurate. In such a situation, the instrument can be considered reliable but not valid. Therefore, for a questionnaire to be scientifically robust and suitable for research application, it should demonstrate both high validity and high reliability [7].

### Determining the Validity of a Questionnaire

The validity of a questionnaire can be evaluated using several approaches depending on the objectives and methodological orientation of the research. Certain forms of validity are more commonly emphasized in quantitative studies, whereas others are particularly relevant in qualitative investigations. Establishing validity is essential to ensure that the questionnaire accurately measures the intended construct and generates meaningful data.

Therefore, this review article aims to discuss the concepts, types, methods, and statistical approaches used in assessing the validity and reliability of questionnaires, with emphasis on their application in health and public health research

### Face Validity

Face validity represents the most basic form of validity assessment and refers to whether the questionnaire appears appropriate and relevant for measuring the intended construct. It involves evaluating the overall presentation, wording, readability, formatting, clarity, and suitability of questionnaire items for the target population. Face validity is primarily based on subjective judgment and may be assessed by researchers, participants, or nonexperts. Although it is considered the least rigorous form of validity, it is particularly useful during the initial stages of questionnaire development to identify confusing or inappropriate items [8].

### Content Validity

Content validity assesses the extent to which a questionnaire adequately covers all dimensions of the construct being investigated. It determines whether the instrument includes a comprehensive representation of the relevant domains and concepts. For example, a questionnaire designed to evaluate cognitive intelligence among school children may include domains such as reading comprehension, mathematical reasoning, logical thinking, and general knowledge. Content validity is generally evaluated by subject experts who examine whether the items are relevant, representative, and sufficient for measuring the intended construct. This process also involves careful consideration of the conceptual basis of the construct being measured. It measures the extent to which experts agree regarding the relevance of questionnaire items.

Experts rate each item using a 4-point Likert scale:

1 = Not relevant

2 = Somewhat relevant

3 = Relevant

4 = Very relevant

Ratings of 3 and 4 are considered relevant.

### Types of CVI

#### 1. Item-Level Content Validity Index (I-CVI)

I-CVI measures the validity of individual items.

Formula:

$$I-CVI = \frac{\text{Number of experts rating item as 3 or 4}}{\text{Total number of experts}}$$

Example: If 4 out of 5 experts rate an item as relevant:

$$I-CVI = 4/5 = 0.80$$

Recommended criteria:

• For  $\leq 5$  experts: I-CVI should be 1.00

• For  $\geq 6$  experts: I-CVI should be at least 0.78



## 2. Scale-Level Content Validity Index (S-CVI)

S-CVI assesses the validity of the overall questionnaire.

Two approaches are commonly used:

### a. S-CVI/UA (Universal Agreement)

Represents the proportion of items receiving relevance ratings from all experts.

### b. S-CVI/Ave (Average)

Represents the average proportion of items rated relevant by experts.

S-CVI/Ave is preferred because universal agreement becomes difficult with increasing numbers of experts. A value  $\geq 0.90$  is generally considered acceptable [9,10].

### Criterion Validity

Criterion validity refers to the extent to which a questionnaire corresponds with an external criterion or established gold standard. It evaluates how effectively the instrument reflects or predicts outcomes measured by another validated method. In quantitative research, criterion validity is commonly assessed by comparing the questionnaire with an already validated tool administered simultaneously. For instance, a shortened quality-of-life questionnaire may be compared with its original

comprehensive version to determine the degree of agreement and correlation between the two instruments [11,12].

### Construct Validity

Construct validity determines whether a questionnaire truly measures the theoretical concept it is intended to assess. It also evaluates whether the questionnaire demonstrates expected relationships with related variables and differentiates itself from unrelated constructs.

Convergent validity examines whether measures that are theoretically related demonstrate strong correlations. For example, a questionnaire assessing satisfaction with a mobile health application would be expected to show positive association with continued app usage and user engagement.

Discriminant validity, also known as divergent validity, assesses the extent to which a questionnaire distinguishes between unrelated constructs. For example, an instrument measuring disease severity should demonstrate changes in scores following effective treatment while remaining unrelated to theoretically irrelevant variables [5,10]

## III. DETERMINING THE RELIABILITY OF A QUESTIONNAIRE

Reliability assessment is an essential step in questionnaire development and is used to determine the consistency, stability, and reproducibility of the instrument. A reliable questionnaire should produce similar results when administered repeatedly under comparable conditions or when evaluated by different observers. Several methods are commonly employed to assess questionnaire reliability.

### Test-Retest Reliability

Test-retest reliability, also known as intrarater reliability, evaluates the stability of a questionnaire over time. In this method, the same questionnaire is administered to the same group of participants on two separate occasions with an appropriate time interval between administrations. The interval should be sufficiently long to minimize recall bias, yet short enough to ensure that the characteristic being measured has not changed significantly during the period.

The suitability of the interval depends on the nature of the construct being assessed. Relatively stable characteristics such as intelligence, personality traits, or long-term interests are less likely to change rapidly, whereas variables such as pain, anxiety, or mood may fluctuate over shorter periods. A reliable questionnaire is expected to demonstrate a strong

correlation between the initial and repeated measurements.

Statistical methods commonly used to evaluate test-retest reliability include Pearson's correlation coefficient and Bland-Altman analysis. High correlation coefficients and minimal mean differences between repeated measurements indicate good stability and consistency of the instrument over time [13].

### Inter-Rater Reliability

Inter-rater reliability assesses the degree of agreement between different observers or assessors using the same research instrument. This form of reliability is particularly important in interviewer-administered questionnaires or observational assessments where responses may be influenced by the evaluator.

To determine inter-rater reliability, multiple assessors independently evaluate the same participants, and the level of agreement between their observations is analyzed. The kappa statistic, including Cohen's kappa and related variants, is commonly used to quantify agreement between raters. A kappa value of 1.0 represents perfect agreement, whereas values below 0.60 generally indicate inadequate agreement between observers. Poor inter-rater reliability suggests that the obtained responses are dependent on the assessor rather than



the instrument itself, thereby reducing the reliability of the questionnaire.

### Internal Consistency

Internal consistency refers to the degree of correlation among items within a questionnaire that are intended to measure the same construct. Strong correlations among related items indicate that the instrument is coherent and reliable. Conversely, weak associations between items may suggest that certain questions are not adequately contributing to the construct being measured [14].

For example, in a questionnaire evaluating cognitive intelligence, items assessing reading ability may correlate more strongly with one another than with items related to mathematical reasoning. Statistical methods such as factor analysis may be employed to identify poorly performing or redundant items within the questionnaire [15].

Internal consistency is most commonly quantified using Cronbach's alpha coefficient. A Cronbach's alpha value of 0.70 or higher is generally considered acceptable and indicates good internal consistency. Values below 0.50 suggest poor inter-item correlation, whereas values above 0.90 may indicate redundancy among questionnaire items [16].

### Importance of Reliability Testing

Reliability testing should be incorporated throughout the questionnaire development process. During pilot testing, assessment of validity and reliability often leads to refinement of questionnaire items, modification of wording, elimination of redundant questions, and improvement in clarity and consistency. Evaluating reliability prior to the final administration of the questionnaire ensures that the instrument accurately and consistently measures the intended construct. Proper planning and systematic psychometric testing are therefore essential for developing a scientifically robust and dependable questionnaire suitable for research applications [10].

### CONCLUSION

Questionnaires remain one of the most widely utilized tools for data collection in health, social, behavioural, and public health research because of their practicality, cost-effectiveness, and ability to obtain information from large populations. However, the scientific value of findings generated through questionnaires depends largely on the psychometric strength of the instrument. Therefore, establishing both validity and reliability is essential during questionnaire development and adaptation. Validity ensures that a questionnaire accurately measures the intended construct, whereas reliability reflects the consistency and stability of the obtained

measurements. Different forms of validity, including face, content, criterion, and construct validity, contribute to evaluating the appropriateness and comprehensiveness of the instrument. Similarly, reliability assessment through methods such as test-retest reliability, inter-rater reliability, and internal consistency analysis ensures reproducibility and precision of responses. The development of a scientifically sound questionnaire requires systematic planning, pilot testing, expert evaluation, statistical assessment, and, when necessary, cultural adaptation and translation. Careful psychometric evaluation not only improves the accuracy and credibility of research findings but also enhances the applicability of questionnaires across diverse populations and research settings. Consequently, researchers must give adequate importance to questionnaire validation and reliability testing to ensure the generation of meaningful, dependable, and high-quality research data.

### REFERENCES

- [1]. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ*. 2004; 328:1312–5.
- [2]. Leung WC. How to design a questionnaire. *BMJ*. 2001; 322:0106187.
- [3]. Ranganathan P, Caduff C. Designing and validating a research questionnaire –Part 1. *Perspect Clin Res*. 2023; 14:152–5.
- [4]. Taherdoost H. Designing a Questionnaire for a Research Paper: a Comprehensive Guide to Design and Develop an Effective Questionnaire. *Asian Journal of Managerial Science*. 2022 Apr 25;11(1):8–16.
- [5]. Ranganathan P, Caduff C, Frampton C. Designing and validating a research questionnaire - Part 2. *Perspectives in Clinical Research* [Internet]. 2024;15(1):42–5.
- [6]. Fendik Saputra. Instrument Reliability Analysis. *Jurnal Indonesia Sosial Sains* [Internet]. 2025 Mar 28;6(3):791–806.
- [7]. Ahmed I, Ishtiaq S. Reliability and validity: Importance in Medical Research. *The Journal of the Pakistan Medical Association* [Internet]. 2021 Oct;71(10):2401–6.
- [8]. Connell J, Carlton J, Grundy A, Taylor Buck E, Keetharuth AD, Ricketts T, et al. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of Life Research*. 2018 Apr 19;27(7):1893–902.
- [9]. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content



- validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*. 2015 Jun 1;4(2):165–78.
- [10]. Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*. 2015;22(4):195–201.
- [11]. Drost EA. Validity and reliability in social science research *Educ Res Perspect*. 2011; 38:105–23
- [12]. Liang Y, Laua PW, Huang YW, Maddison R, Baranowski T. Validity and reliability of questionnaires measuring physical activity self-efficacy, enjoyment, social support among Hong Kong Chinese children *Prev Med Rep*. 2014;1:48–52
- [13]. Pedisic Z, Bennie JA, Timperio AF, Crawford DA, Dunstan DW, Bauman AE, et al Workplace sitting breaks questionnaire (SITBRQ): An assessment of concurrent validity and test-retest reliability *BMC Public Health*. 2014; 14:1249
- [14]. Anastasiadou SD. Reliability and validity testing of a new scale for measuring attitudes and toward learning statistics with technology *Acta Didactica Napocensia*. 2011; 4:1–1046.
15. Cronbach LJ. Coefficient alpha and the internal structure of tests *Psychometrika*. 1951; 16:297–334
- [16]. Tavakol M, Dennick R. Making sense of Cronbach's alpha *Int J Med Educ*. 2011;2:53–5