# Validation of questionnaires for the measurement of health variables: Fundamental concepts

Natalia Riva[a] , Diego Grandi[a] , Benjamín Cruzat[a] , Ruben Alvarado[b*]

[a]Estudiante de Medicina, Escuela de Medicina, Facultad de Medicina, Universidad de Valparaíso. Valparaíso, Chile; [b]Departamento de Salud Pública, Escuela de Medicina, Facultad de Medicina, Universidad de Valparaíso. Valparaíso, Chile

**ABSTRACT**

In clinical practice and population health, it is common to use questionnaires to assess conditions or variables that are not directly observable. However, the construction and validation of these instruments or questionnaires are often poorly understood. This narrative review aims to summarise in a general way the process of construction and validation of these questionnaires in order to have a better understanding of this process, the aspects that are evaluated, and the best way to use them. The validation of questionnaires corresponds to a process of analysis of the questionnaire, aiming to measure a latent variable and its dimensions, which cannot be observed directly. A latent variable can be inferred through a set of specific attributes that are part of it, such as the items of a questionnaire, which are observable. Through a narrative review, this article addresses the fundamental concepts of questionnaire or test validation, latent variables or constructs, reliability and validity studies, and the factors that theoretically affect the latter two characteristics. Examples of these concepts are presented in the tex

**KEYWORDS** Reliability and Validity, Surveys and Questionnaire, Epidemiology

## INTRODUCTION

Questionnaire validation corresponds to a questionnaire analysis process that measures constructs or latent variables and possible dimensions that cannot be observed directly. The questionnaire comprises several questions or items through which the measurement is sought. Subjects' response to the items is how we measure a variable that cannot be directly observed. As Carmines and Zeller point out, it is "the process of linking abstract concepts with empirical indicators through an explicit and organized plan to qualify the available data, in terms of the concept the researcher has in mind" [1].

An important and distinctive aspect of psychometrics is how the measurement concept is understood. In this case, it is an associative action (of a numerical value with a given response), not just assigning numerical values to specific items.

This literature review is the thirteenth installment of a methodological series of narrative reviews on general biostatistics and clinical epidemiology topics, which explore and summarize published articles in user-friendly language available in major databases and specialized reference texts. The series is oriented to the training of undergraduate and graduate students. It is carried out by the Chair of Evidence-Based Medicine of the School of Medicine of the University of Valparaiso, Chile, in collaboration with the Research Department of the University Institute of the Italian Hospital of Buenos Aires, Argentina, and the Evidence Center of the Pontificia Universidad Católica de Chile. This manuscript aims to describe the process of instrument validation and the elements that comprise it.

Regarding the definition of latent variable [2], it refers to a condition that is not directly observable, but that can be inferred through a set of specific observable attributes that are part of it. Examples of latent variables are self-esteem, social support, satisfaction with services delivered, or quality of life. Many conditions in medicine and public health correspond to latent variables that require investigation. Consequently, we can measure a latent variable using people's answers to questionnaire items.

For example, in the evaluation of the psychomotor development of an infant, we can consider as a non-observable attribute the spatial skills and as an observable attribute the action of assembling towers of blocks during an evaluation of psychomotor development, thus identifying the visuospatial skills according to the height of the blocks that the infant manages to assemble.

**MAIN MESSAGES**

- Latent variables cannot be directly observed and must be measured through questionnaires demonstrating good validity and reliability.
- The reliability of a questionnaire refers to the degree to which it measures without errors.
- The validity of a questionnaire corresponds to the degree to which it measures the latent variable and is usually evidenced in three different aspects: the conceptual concordance of its content, the concordance against an external criterion, and the concordance with the theoretical construct.

## METHODOLOGICAL ASPECTS OF VALIDATING MEASUREMENT INSTRUMENTS

### Requirements for a good measurement

In order to carry out a good measurement, it is necessary to select the appropriate questionnaire or test, which must reliably represent the variable being studied. It will, therefore, have to meet three criteria:

*Reliability*

The degree to which the questionnaire produces consistent results in different measurements, either by its various items, by different interviewers, or at different moments.

*Validity*

The degree to which the questionnaire measures the variable we intend to evaluate.

*Objectivity*

Refers to the degree to which the questionnaire allows to assess the characteristics of an object as it is, avoiding the subjective aspects of the person who administers, classifies, and interprets and reducing as much as possible the biases that this may entail.

## RELIABILITY: CONCEPT AND DEFINITION

The concept of reliability, consistency, or dependability of a test is related to the random measurement errors present in the scores obtained from its application. In other words, it is the ability to make an error-free measurement [3].

For this, we must introduce the concept of internal consistency [4], which is the degree to which each of the instrument's parts is composed and equivalent to the rest [2]. Therefore, within this concept, we will have three dimensions of reliability that can be studied:

1. Consistency of its set of items.
2. Consistency over time, stability, or mediation.
3. Consistency in application by different interviewers.

There are two elements to consider when assessing reliability. First, select and use the best questionnaire or test for the topic under study. Secondly, the level of error that the measurement given by the questionnaire may have. Thus, our objective will be the construction and use of tests that allow us to significantly reduce the error, bearing in mind that when carrying out a measurement, we will always be in the presence of variations due to chance, which cannot be eliminated.

### Reliability: intra-test reliability or internal consistency

Following the definition mentioned above, internal consistency can be calculated through various statistical tests, depending on the characteristics of the responses to its items. One of the most commonly used tests is Cronbach's α, which expresses the degree of co-variance of items within a test or part of a test. It is the appropriate test for tests whose items have several response alternatives.

The Cronbach's α formula [5] is presented in Figure 1.

Since this is a covariate, the best internal consistency is obtained with values close to 1. In general, the accepted ranges are as follows:

1. Values below 0.5 points are unacceptable.
2. Values between 0.5 and 0.6 points have poor consistency.
3. Values between 0.6 and 0.7 points are of questionable consistency.
4. Values between 0.7 and 0.8 points are of acceptable consistency.
5. Values between 0.8 and 0.9 points have a good consistency.
6. Values higher than 0.9 points have excellent consistency [6].

Specifically, for psychological assessments, there is a consensus on values between 0.65 and 0.8, which are considered adequate [7].

For example, in the validation of the Warwick-Edinburgh mental well-being instrument in the Chilean population, carried out by Carvajal et al. [8], Cronbach's α was used to evaluate the internal consistency of the 14 items, with a value of 0.875 (which is good) and without the need to eliminate any item to improve this indicator.

### Reliability: test-retest reliability

This method is based on the existence of a temporal stability of the construct being assessed through the questionnaire. If two measurements are carried out separated by a reasonable time, the results in both tests should not vary significantly. It allows us to use the same test on both occasions [3], giving us

the advantage of not having to construct a different questionnaire to measure the same latent variable [2].

It can be calculated using some correlation coefficient between the first and second measurements (e.g., Pearson's r or Spearman's Rho) [9], where a value close to 1 would indicate a positive correlation between the results of both applications and a value close to 0 would indicate that there is no correlation between these two applications over time.

One of the essential factors to consider is the time lapse between the first and the second application of the test, as this must be consistent with a time in which the latent variable is expected not to change significantly. Another relevant condition is that there is no alteration in the number of subjects tested, as this could introduce a selection bias, decreasing the reliability of the measurement.

In the same previous example by Carvajal et al. [8], the test-retest assessment was done two weeks apart and in a sub-sample of 50 people (22.7% of the total sample), with a good value for Spearman's Rho correlation (0.556) and p < 0.001.

### Reliability: inter-rater reliability or Cohen's κ

In the clinical context, multiple entities often assess the same subject, so there must be concordance in the classifications from two or more test administrators [2].

As in the other reliability analyses described above, several tests can be used depending on the type of item response. One of the most commonly used tests is the kappa coefficient (κ), which is helpful for items with nominal scales. For example, we have two physicians, one specialist and one non-specialist, who must determine whether a group of patients has preserved or reduced cardiac ejection fraction (nominal classification). As this test will depend on the performer's experience, we may have a difference with the results obtained, so to measure the reliability of this assessment, the κ coefficient can be applied. The modified κ coefficient or the intraclass correlation coefficient can be used when dealing with items with ordinal or interval scales. Using the same example above, if we wanted to assess the percentage of reduced ejection fraction (ordinal classification), we could also have differences with the assessors.

**Figure 1.** Cronbach's α formula for determining intra-test reliability.

$$\alpha = \frac{K}{K-1}\left(\frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

*a: Cronbach's alpha.*
$\kappa$: Number of questions or items considered.
$\sigma_{Y_i}^2$: Item variance.
$\sigma_X^2$: Variance of observed values.
Source: Prepared by the authors of this study based on the following source [4].

In this case, assessing this using the modified κ coefficient would be sensible.

The formula for calculating Cohen's κ coefficient [10] can be seen in Figure 2.

Values close to 1 will indicate high consistency between the results obtained by different interviewers. On the other hand, values close to 0 indicate that there is low consistency between them. In other words, it means that the results obtained by different interviewers are determined more by other factors than by the questionnaire itself [10].

## VALIDITY: CONCEPT AND DEFINITION

The degree of validity of a questionnaire or test is one of its fundamental characteristics. As mentioned above, validity refers to whether the test "measures what it is intended to measure", i.e., whether it assesses that variable and not something else [11,12]. Thus, a questionnaire that intends to measure the level of leadership measures that and not the level of autonomy, for example.

In order to establish that a test or questionnaire has an adequate level of validity, evidence of three aspects of this validity is sought: content validity, criterion validity, and construct validity [10,13].

### Validity: content

The first dimension of validity is content validity. Content validity is assessed by the set of items that make up the questionnaire. These are expected to be a representative sample of the construct or latent variable being studied. That is, the set of items includes all the aspects involved in the concept of that latent variable, and there are no items assessing aspects that are not included in that latent variable.

To measure knowledge about a topic, we should consider including all the key aspects of what is being measured [14]. For example, when measuring the presence of a depressive disorder, a questionnaire that asks only about affective state characteristics (despondency, apathy, etc.) but not about cognitive symptoms (pessimism, low self-esteem, etc.) or physical symptoms (changes in appetite or sleep), would not have sufficient content validity. On the other hand, if it incorporates symptoms that are not typical of depressive disorders but are often associated with them (such as alcohol abuse or panic attacks), its content validity would also be reduced.

A questionnaire with good content validity must adequately measure all the main dimensions that form part of the variable under study [15].

### Validity: criterion-related

Criterion validity is obtained when we compare the results of the questionnaire we are testing with those of another test that measures the same or a related construct and has previously been shown to have good validity [16].

Figure 2. Calculation of inter-rater reliability on nominal scales through Cohen's κ.

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

*κ:* Cohen Kappa.
$\rho_o$o: *proportion of individuals consistently ranked.*
$\rho_e$: proportion of individuals random ranked.
Source: Prepared by the authors of this study based on the following source [4].

If the criterion we compare to our test can be applied and assessed in the present, we speak of "concurrent validity".

For example, Silva et al. (2012) adapted and validated the Eating Disorders Diagnostic Scale (EDDS) [17] into Spanish, a questionnaire designed for the diagnosis of eating disorders. Selecting a group of patients and a control group, they applied the EDDS and the structured psychiatric interview (CIDI, Composite International Diagnostic Interview), which allows for assessing the presence of mental disorders in the respondent, according to the classifications of the Diagnostic and Statistical Manual of Psychiatric Disorders, IV edition (DSM-IV) and the International Classification of Diseases, 10th edition (ICD-10). Using the ICD as a gold standard, they created contingency tables and found a moderate to high correlation between the results of both tests for diagnosing the presence of an eating disorder.

Now, when the criterion against which we compare our test is applied in the future, we are talking about "predictive validity". Predictive validity is assessed with statistical tests that depend on the type of variables we are working. It could be a t-test when comparing averages [18], a Pearson or Spearman correlation when dealing with continuous variables [19], or a concordance test when dealing with two dichotomous variables [19,20]. An example of this type of validity is the study by Larzelere et al. (1996), where the predictive validity of the Suicide Probability Scale questionnaire was assessed [21]. In this study, suicide attempts, suicidal verbalization, and self-harming behaviors following the application of this questionnaire were analyzed, which was carried out at the time of admission to a shelter for children at risk.

### Validity: construct

Construct validity refers to how well a test measures and represents the theoretical concept (construct or latent variable) being assessed. It involves checking the theory against the empirical evidence obtained by applying the test to subjects. For example, if the theory states that the construct under study has two dimensions and the questionnaire used has items that measure these two dimensions, it would be expected that there is a high correlation between the items within each dimension, and at the same time that there is a low correlation between the items measuring the two different dimensions [22].

For the above, the most commonly used test is factor analysis [23]. This test seeks to identify patterns of association between variables or items to gather into "factors" most correlated with each other. In this way, a questionnaire shows good construct validity when, in the factor analysis, the items that make up a theoretical dimension are grouped into the same factor.

For example, Alvarado et al. (2015) evaluated the Edinburgh questionnaire [24] to identify depressive disorders during pregnancy. The exploratory factor analysis found that the 12 items that make up this test were gathered into a single factor, as theoretically expected.

## FACTORS AFFECTING RELIABILITY AND VALIDITY

As mentioned above, "random error" and "measurement bias" are closely related to reliability and validity [25]. Bias corresponds to a systematic tendency to underestimate or overestimate the estimator of interest [25] caused by a deficiency in the design or execution of a study [26], negatively affecting the validity of the questionnaire. On the other hand, random error corresponds to variations explained by chance [27] and cannot be eliminated (although it can be minimized), negatively affecting the test's reliability. There are three main factors [28] associated with the degree of random error and how it might affect our measurement:

1. The degree of individual and inter-individual variability.
2. The size of the sample.
3. The magnitude of the differences found.

With this in mind, measures can be taken to reduce random error by increasing the sample size or establishing measurement targets with larger differences in magnitude. An example of random error in measurements can be seen in estimating fetal weight through ultrasound by biometric method, where differences of ±30 grams concerning fetal weight have been reported [29]. Other factors affect the final validity and reliability of a questionnaire, and these are mentioned below, organized according to the stage at which they may appear in the process of test construction and validation:

### Construction of the questionnaire

At this stage, we must try to ensure that the test is constructed in an ideal way for what we are trying to measure. **Improvisation** [30], little research into the subject of interest, and little experience in creating collection questionnaires leads to poor construction, which can increase random and systematic error. Another essential point is the use of **questionnaires validated abroad** [30], which **may not be contextualized** for the research in our interest group regardless of being translated and adapted to the local language [31]. On the other hand, the **standardization process** allows us to reduce the measurement bias of the study by using the same questionnaires and ways of measuring the variable of interest in all participants. Finally, the **mechanical**

**Figure 3.** Infographic validation of instruments measuring latent variables or constructs.
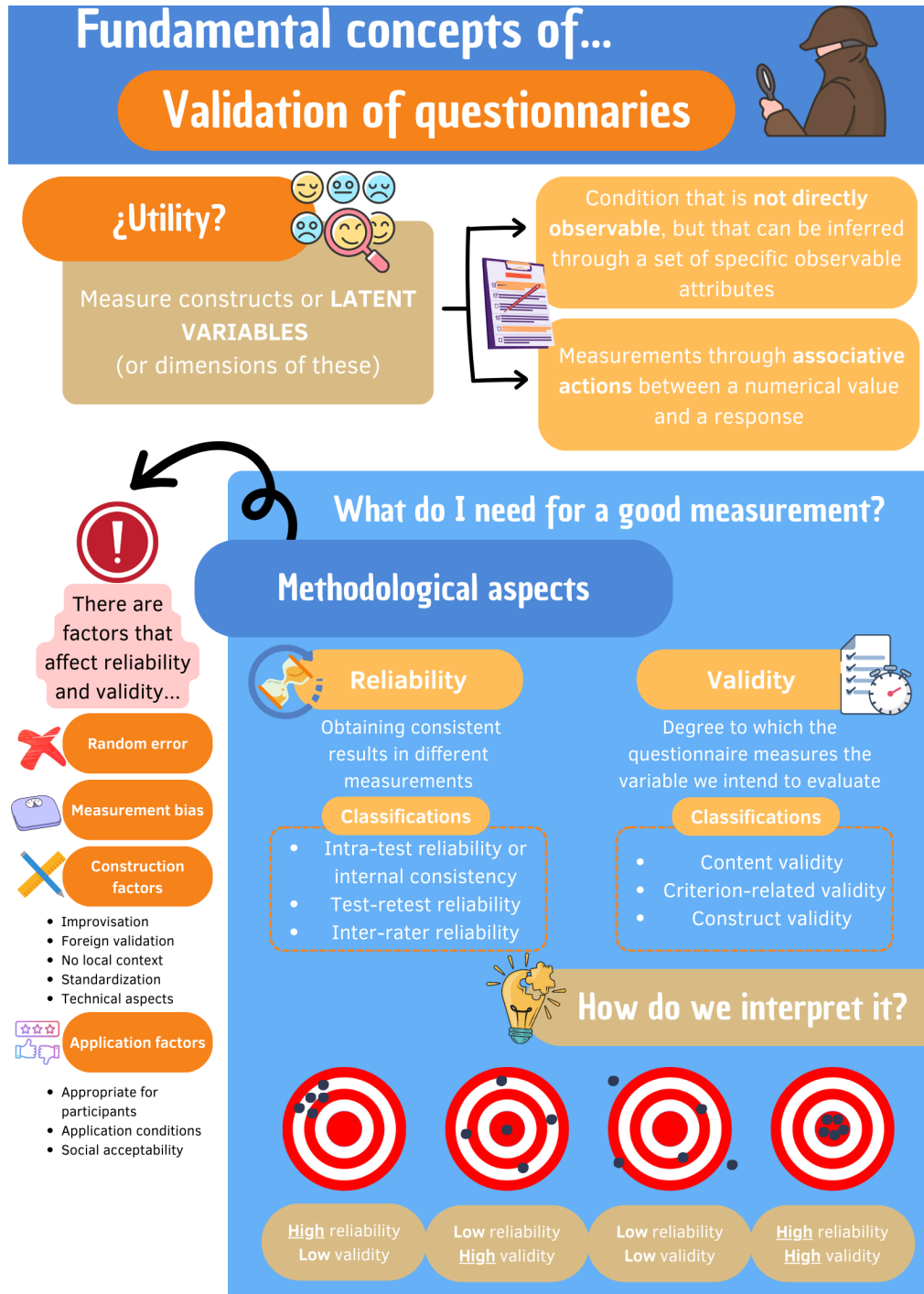


Source: Prepared by the authors of this study.

aspects of the test construction [30] should be mentioned. For example, it is not easy to understand if the instructions are unclear or use many technical terms.

**Application of the questionnaire**

At this point, it should be considered that the test should be appropriate [30] for the person to whom it is being administered (e.g., the use of reading materials in subjects who cannot read or not considering important differences based on gender). In addition, the application conditions must be appropriate,

both contextually and for the individual (e.g., providing a quiet space that allows for concentration without pressure). Another element to highlight is the participants' style [30,32], directing responses toward what is socially acceptable and omitting what is undesirable. This is known as social acceptability bias or social complacency bias. For example, a person responds that their child's meals do not usually include the consumption of sugary drinks, tending to respond to what is assumed to be socially desirable.

## FINAL CONSIDERATIONS

Questionnaire validation is a key process in measuring latent variables or their dimensions, which cannot be directly observed. Questionnaires must have high reliability and good validity in terms of error and bias to ensure that measurement is as accurate as possible. Reliability refers to the degree to which a questionnaire generates accurate and consistent results, while validity relates to the degree to which the test measures the variable it is intended to assess.

Reliability can be assessed by internal consistency, test-retest, and inter-rater reliability tests. On the other hand, validity is composed of three dimensions: content, criterion, and construct. Content validity refers to the representativeness and relevance of the questionnaire items, while criterion validity involves comparing the questionnaire results with other questionnaires or already validated criteria. Finally, construct validity relates to evidence that the questionnaire measures the expected theoretical construct.

Finally, during the investigation of latent variables, the existence of factors that may affect the validity and reliability of the questionnaires during their construction, choice, and application must be considered. By following these methodological principles, greater accuracy of the results is ensured, and analyses and conclusions can be made with low error rates (Figure 3).

A good example of the usefulness of questionnaires is the incorporation of this type of instrument to improve the detection of clinical conditions that usually go unnoticed, such as depressive conditions in pregnancy. Alvarado et al. [24] validated a questionnaire that showed good validity and reliability indicators for detecting this problem, which is why it is currently incorporated as a tool in pregnancy check-ups.

## REFERENCES

1.  Carmines EG, Zeller RA. In: Reliability and Validity Assessment [Internet]. SAGE Publications; 1979. https://play.google.com/store/books/details?id=o5x1AwAAQBAJ

2.  Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. Psychol Rev. 2003;110: 203–219. http://dx.doi.org/10.1037/0033-295X.110.2.203 https://doi.org/10.1037/0033-295X.110.2.203

3.  Meneses J, Barrios M, Lozano LM, Bonillo A, Turbany J, Cosculluela A, et al. Psicometría. Editorial UOC. 2014. https://play.google.com/store/books/details?id=2JxuBAAAQBAJ

4.  Manterola C, Grande L, Otzen T, García N, Salazar P, Quiroz G. Reliability, precision or reproducibility of the measurements. Methods of assessment, utility and applications in clinical practice. Rev Chilena Infectol. 2018;35: 680–8. http://dx.doi.org/10.4067/S0716-10182018000600680 https://doi.org/10.4067/S0716-10182018000600680

5.  Sijtsma K. In: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha [Internet]. http://dx.doi.org/10.1007/s11336-008-9101-0https://doi.org/10.1007/s11336-008-9101-0

6.  Toro R, Peña-Sarmiento M, Avendaño-Prieto BL, Mejía-Vélez S, Bernal-Torres A. Análisis Empírico del Coeficiente Alfa de Cronbach según Opciones de Respuesta, Muestra y Observaciones Atípicas. REV IBEROAM DIAGN EV. 2022;63: 17. https://www.redalyc.org/journal/4596/459671926003/459671926003.pdf https://doi.org/10.21865/RIDEP63.2.02

7.  Komorita SS, Graham WK. Number of Scale Points and the Reliability of Scales. Educational and Psychological Measurement. 1965;25: 987–995. https://journals.sagepub.com/doi/10.1177/001316446502500404 https://doi.org/10.1177/001316446502500404

8.  Carvajal D, Aboaja A, Alvarado R. Validación de la Escala de bienestar mental de Wareick-Edinburgo, en Chile. Rev Salud Pública (Córdoba). 2015;19: 13–21. https://revistas.unc.edu.ar/index.php/RSD/article/view/11822

9.  Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18: 91–93. http://dx.doi.org/10.1016/j.tjem.2018.08.001 https://doi.org/10.1016/j.tjem.2018.08.001

10. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med [Internet]. 2012;22: 276–82. http://dx.doi.org/10.1016/j.jocd.2012.03.005 https://doi.org/10.1016/j.jocd.2012.03.005

11. Sage Journals. 3 Oct 2023. https://journals.sagepub.com/action/cookieAbsent

12. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25: 508–18. http://dx.doi.org/10.1177/

0193945903252998 https://doi.org/10.1177/
0193945903252998

13. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist. 1995;50: 741–749. https://psycnet.apa.org/fulltext/1996-10004-001.pdf https://doi.org/10.1037/0003-066X.50.9.741

14. American Educational Research Association. Standards for Educational and Psychological Testing. 2014. https://play.google.com/store/books/details?id=cll_mAEACAAJ

15. Schuwirth LWT, van der Vleuten CPM. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach. 2011;33: 783–97. http://dx.doi.org/10.3109/0142159X.2011.611022 https://doi.org/10.3109/0142159X.2011.611022

16. Kanya L, Sanghera S, Lewin A, Fox-Rushby J. The criterion validity of willingness to pay methods: A systematic review and meta-analysis of the evidence. Soc Sci Med. 2019;232: 238–261. http://dx.doi.org/10.1016/j.socscimed.2019.04.015 https://doi.org/10.1016/j.socscimed.2019.04.015

17. Silva JR, Behar R, Cordella P, Ortiz M, Jaramillo K, Alvarado R, et al. Validation of the Spanish version of the Eating Disorders Diagnostic Scale. Rev Med Chil. 2012;140: 1562–70. http://dx.doi.org/10.4067/S0034-98872012001200007 https://doi.org/10.4067/S0034-98872012001200007

18. Wadhwa RR, Test R. T Test. In: StatPearls [Internet] [Internet]. 2023. https://pubmed.ncbi.nlm.nih.gov/31971709/

19. Beath A, Jones MP. Guided by the research design: choosing the right statistical test. Med J Aust. 2018;208: 163–165. https://pubmed.ncbi.nlm.nih.gov/29490219/ https://doi.org/10.5694/mja17.00422

20. Cortés-Reyes É, Rubio-Romero JA, Gaitán-Duarte H. Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. Rev Colomb Obstet Ginecol. 2010;61: 247–255. http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0034-74342010000300009&lng=en&nrm=iso&tlng=es https://doi.org/10.18597/rcog.271

21. Larzelere RE, Smith GL, Batenhorst LM, Kelly DB. Predictive validity of the suicide probability scale among adolescents in group home treatment. J Am Acad Child Adolesc Psychiatry. 1996;35: 166–72. https://pubmed.ncbi.nlm.nih.gov/8720626/ https://doi.org/10.1097/00004583-199602000-00009

22. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955;52: 281–302. https://doi.org/10.1037/h0040957

23. Cuadras C. In: Nuevos métodos de análisis multivariante [Internet]. CMC Editions; 2014. https://gc.scalahed.com/recursos/files/r161r/w24899w/Semana5/METODOS_S5.pdf

24. Alvarado R, Jadresic E, Guajardo V, Rojas G. First validation of A Spanish-translated version of the Edinburgh postnatal depression scale (EPDS) for use in pregnant women. A Chilean study. Arch Womens Ment Health. 2015;18: 607–12. http://dx.doi.org/10.1007/s00737-014-0466-z https://doi.org/10.1007/s00737-014-0466-z

25. Gempp Fuentealba R. El error estándar de medida y la puntuación verdadera de los tests psicológicos: Algunas recomendaciones prácticas. Terapia [Internet]. 2006;24: 117–29. https://www.redalyc.org/articulo.oa?id=78524201

26. Barraza F, Arancibia M, Madrid E, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Random error and systematic error. Medwave. 2019;19. https://doi.org/10.5867/medwave.2019.07.7687

27. Vetter TR, Mascha EJ. Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My! Anesth Analg. 2017;125: 1042–1048. https://journals.lww.com/anesthesia-analgesia/abstract/2017/09000/bias,_confounding,_and_interaction__lions_and.46.aspx https://doi.org/10.1213/ANE.0000000000002332

28. Bowling A. Research Methods In Health: Investigating Health And Health Services. McGraw-Hill Education (UK); 2014. https://play.google.com/store/books/details?id=6lOLBgAAQBAJ

29. Ferreiro RM, Valdés Amador L. Eficacia de distintas fórmulas ecográficas en la estimación del peso fetal a término. Rev Cubana Obstet Ginecol. 2010. http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0138-600X2010000400003&lng=es&nrm=iso&tlng=es

30. Hernández Sampieri R, Fernández Collado C, Bautista Lucio P. Metodología de la investigación. McGraw-Hill. 2014.

31. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. Salud Publica Mex. 2013;55: 57–66. http://dx.doi.org/10.1590/s0036-36342013000100009 https://doi.org/10.1590/s0036-36342013000100009

32. Foddy W. Constructing Questions for Interviews and Questionnaires. Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research. Cambridge University Press; 1993. https://www.cambridge.org/core/product/identifier/9780511518201/type/book https://doi.org/10.1017/CBO9780511518201

# Validación de cuestionarios para la medición de variables en salud: conceptos fundamentales

# Validation of questionnaires for the measurement of health variables: Fundamental concepts

**ABSTRACT**

Dentro de la práctica clínica, así como en la salud poblacional, es habitual utilizar cuestionarios que permiten evaluar condiciones o variables que no son directamente observables. No obstante, la construcción y validación de estos instrumentos o cuestionarios suele ser poco conocida. El objetivo de esta revisión narrativa es sintetizar de manera general el proceso de construcción y validación de estos cuestionarios, para así tener una mejor comprensión de este proceso, de los aspectos que se evalúan y de la mejor forma de utilizarlos. La validación de cuestionarios corresponde a un proceso de análisis de este, cuya finalidad es medir una variable latente o constructo, así como sus dimensiones, las que no pueden ser observadas directamente. Una variable latente puede ser inferida a través de un conjunto de atributos específicos que forman parte de ella, como los ítems de un cuestionario y que sí son observables. En este artículo se abordan de manera teórica los conceptos fundamentales de validación de cuestionarios o test, variables latentes o constructos, estudio de la confiabilidad y de la validez, así como los factores que afectan a estas dos últimas características, a través de una revisión narrativa. En el texto, se presentan ejemplos sobre estos conceptos.