

## Objetividad, validez y confiabilidad: atributos científicos de los instrumentos de medición

### Objectivity, validity and reliability: scientific attributes of measuring instruments

María L. Bautista-Díaz <sup>a</sup>, Karina Franco-Paredes <sup>b</sup>, Hortensia Hickman Rodríguez <sup>c</sup>

---

#### Abstract:

Psychological tests are instruments frequently used to measure psychological constructs. Those who use these tests should know their scientific attributes of them. Thus, the present diagram aimed to expose the objectivity, validity, and reliability as scientific attributes of psychological measurement instruments. The generation of evidence of the metric quality of a test begins during the construction process and it is important to analyze the test when it is used in contexts or countries other than where it was created. Likewise, it is necessary to continually generate evidence regarding the validity and reliability of test scores to ensure that they meet the psychometric standards required to be used.

#### Keywords:

*Objectivity, validity, reliability, psychological tests*

---

#### Resumen:

Las pruebas psicológicas son instrumentos que se utilizan frecuentemente para medir constructos psicológicos. Quienes utilizan estas pruebas es importante que conozcan los atributos científicos de las mismas. Por ello, el objetivo del presente diagrama fue exponer la objetividad, validez y confiabilidad como atributos científicos de los instrumentos psicológicos de medición. La generación de la evidencia de la calidad métrica de una prueba inicia desde el proceso de construcción y es importante que la prueba sea analizada cuando se utiliza en contextos o países diferentes a donde fue creada. Asimismo, es necesario que continuamente se genere evidencia respecto a la validez y confiabilidad de las puntuaciones de una prueba para garantizar que cumplen con los estándares psicométricos requeridos para ser utilizada.

#### Palabras Clave:

*Objetividad, validez, confiabilidad, pruebas psicológicas*

---

### Introducción

En la actividad científica de las ciencias sociales en general y de la psicología en específico, desde la formación de pregrado, posgrado y la práctica profesional, un paso importante en la investigación

cuantitativa después de que se ha determinado un diseño de investigación y las unidades de observación, es la selección de los medios de recolección de datos. Estos pueden ser entrevistas estructuradas, autoinformes (cuestionarios, inventarios o escalas) o bitácoras, su elección estará en función de los objetivos del

---

<sup>a</sup> Autor de Correspondencia, Universidad Nacional Autónoma México, <https://orcid.org/0000-0003-1154-1737>, Email: [leticia.bautista@iztacala.unam.mx](mailto:leticia.bautista@iztacala.unam.mx)

<sup>b</sup> Universidad de Guadalajara, <https://orcid.org/0000-0002-5899-3071>, Email: [karina.franco@cusur.udg.mx](mailto:karina.franco@cusur.udg.mx)

<sup>c</sup> Universidad Nacional Autónoma México, <https://orcid.org/0000-0002-4025-9485>, Email: [hortensiahickman@gmail.com](mailto:hortensiahickman@gmail.com)

profesional/investigador o del tipo de variables o constructos a evaluar. <sup>[1]</sup>

Los autoinformes son los instrumentos de evaluación más utilizados por profesionales de la psicología <sup>[2]</sup> y por otros profesionales de las ciencias sociales ya que, permiten ahorrar en recursos humanos, económicos y de tiempo, en virtud de su practicidad para realizar evaluaciones grupales estandarizadas (mismas condiciones) en menor tiempo y con menor número de aplicadores o evaluadores. Sin embargo, para que los datos obtenidos de los cuestionarios sean útiles científicamente se requiere que estos posean propiedades psicométricas adecuadas; entendiendo por propiedades psicométricas como los atributos científicos que surgen de una teoría y que son susceptibles de demostrarse. <sup>[3]</sup> Por tanto, el objetivo del presente diagrama fue exponer la objetividad, validez y confiabilidad como atributos científicos de los instrumentos de medición.

## Desarrollo

En el caso concreto de la Teoría Clásica de los Tests <sup>[4,5]</sup> un postulado importante es analizar un cuestionario con la mayor precisión científica para valorar su calidad, aunque sería imposible evitar el error en la medición, se intenta minimizarlo, ya que, a partir de las puntuaciones de dichos cuestionarios se hacen inferencias y se toman decisiones valiosas, por ello, se deben evaluar y demostrar sus propiedades psicométricas.

La psicometría permite operacionalizar, cuantificar y probar teorías en un medio de evaluación de autoinforme, es decir, posibilita medir científicamente lo no observable (lo psicológico), por esta razón, es que a algunos instrumentos de evaluación se les conoce como pruebas psicométricas. La historia de la psicometría es joven, data de mediados del siglo pasado cuando las dos guerras mundiales colocaron las condiciones idóneas para realizar evaluaciones grupales o en masa. Sin embargo, la práctica de aplicar cuestionarios de manera rutinaria sin un propósito concreto llevó a regular dicha actividad, entonces se exigió que cualquier instrumento de evaluación debería poseer tres atributos: objetividad, validez y confiabilidad. <sup>[6,7]</sup>

### Objetividad

En los instrumentos de medición "la objetividad" ha generado crítica <sup>[1]</sup>, ya que resulta contradictorio medir *objetivamente* lo no observable, por ello, se trata de garantizarla en la mayor medida. Se ha documentado que la objetividad tiene al menos cinco características: especificidad, neutralidad, independencia, imparcialidad

e impersonalidad. <sup>[1]</sup> Por ejemplo, si se dice que se va a estudiar la mente, primero se debe operacionalizar el constructo y después, avalar que el medio de evaluación se corresponda con los indicadores contenidos en el instrumento. Es claro entonces que, se trata de un proceso arduo que antecede al atributo de validez y confiabilidad, por tanto, la objetividad no es sinónimo de validez y confiabilidad, sino un atributo primordial exigible en cualquier instrumento de medición.

### Validez

Una vez garantizada la objetividad, es imperante demostrar la validez, la cual se refiere al grado en que la evidencia y la teoría apoyan la interpretación de las puntuaciones de los tests, para el propósito con el que son utilizados <sup>[8]</sup> Como se observa en la Figura 1, la validez de un instrumento va más allá de realizar un análisis, se requiere generar evidencia respecto a la validez relacionada con el contenido, con el criterio y con el constructo. <sup>[9]</sup> Concretamente, la validez relacionada con el contenido alude a garantizar que los indicadores (reactivos) de una prueba representan una porción de un universo posible, esto es sencillo cuando se tienen indicadores objetivos tal es el caso de la evaluación de las habilidades matemáticas, pero en el caso de variables psicológicas o constructos (entidades no tangibles), se opta por llevar a cabo un jueceo de expertos quienes valoran si un conjunto de reactivos es idóneo para medir un constructo específico. <sup>[10,11]</sup>

No obstante, además del jueceo de expertos se puede solicitar a una porción de la población blanco para que emita su juicio sobre los indicadores de una prueba. A esta muestra de la población se identifica como *jueces no expertos*, ya que, sin conocer conceptual o metodológicamente la teoría psicométrica, retroalimentan la prueba. De este modo, con los datos proporcionados por los jueces expertos y los no expertos se puede derivar un índice de validez de contenido, que en palabras sencillas es el promedio del acuerdo entre los jueces. Se espera un acuerdo mayor al 80% para determinar que una prueba posee adecuada validez de contenido. <sup>[12]</sup> En este caso se sugiere describir todo el proceso llevado a cabo en la construcción de una prueba, desde la definición del dominio, diseño del banco de reactivos, así como su calidad, representación, pertinencia y congruencia, acompañado de su respectivo índice de validez de contenido, seguido de un piloteo de la versión preliminar ya que, todo lo anterior será evidencia de la validez relacionada con el contenido. <sup>[13]</sup>

En cuanto a la validez relacionada con el criterio (Ver Figura 1), puede ser de tipo concurrente o predictiva. En

el caso de la primera se debe utilizar -paralelamente- como criterio otra prueba, un diagnóstico o una tarea que midan el mismo constructo. Posteriormente, se calcula un coeficiente de correlación a los datos derivados de la prueba y del criterio externo concurrente, se espera un grado de asociación de moderado a elevado con una dirección positiva, ya que se evalúa el mismo constructo. En la validez relacionada con el criterio de tipo predictiva, el criterio se evalúa a futuro, por ejemplo, en una empresa se evalúan las competencias de ventas en una serie de candidatos para su contratación y en un lapso posterior se evalúan sus ventas reales. Una vez hecho esto, se calcula un coeficiente de correlación o de regresión a los datos derivados de la prueba y el criterio externo predictor, demostrando, según sea el caso, que la prueba posee validez relacionada con el criterio. [6,9]

Otro procedimiento consiste en analizar los cambios relacionados con el desarrollo. Si un constructo se comporta diferencialmente de acuerdo con las etapas del desarrollo, se deben calcular pruebas estadísticas de comparación, en este caso diferentes grupos etarios para confirmar que obtienen diferentes puntuaciones en el instrumento. [6] Este mismo procedimiento se aplica para generar la evidencia de diferencias grupales consistentes con la teoría, puede ser considerado el sexo, el estatus socioeconómico, la procedencia (rural-urbana) o nivel académico, entre otras que confirmarán o no las teorías.

Demostrar que el instrumento es sensible a identificar el efecto de una intervención *ex profeso* es un procedimiento que se conoce como efectos de la intervención. En este caso con pruebas estadísticas para grupos relacionados (diseños pretest-postest) se analiza si el instrumento identifica cambios posteriores a una intervención. Otros procedimientos como validez convergente y divergente deben utilizar pruebas estadísticas de correlación. En la primera (convergente) se selecciona una prueba en la cual sus indicadores tengan convergencia con el constructo a validar, se espera una correlación al menos moderada con el constructo. Por ejemplo, si se tiene una prueba que evalúa el constructo de bienestar psicológico y otra prueba que evalúa calidad de vida, al tratarse de dos constructos que teóricamente están relacionados se tiene que demostrar una correlación entre ambos; en el caso de la de tipo divergente, se esperaría asociaciones negativas dado que teóricamente los constructos no están correlacionados. [3,6,9]

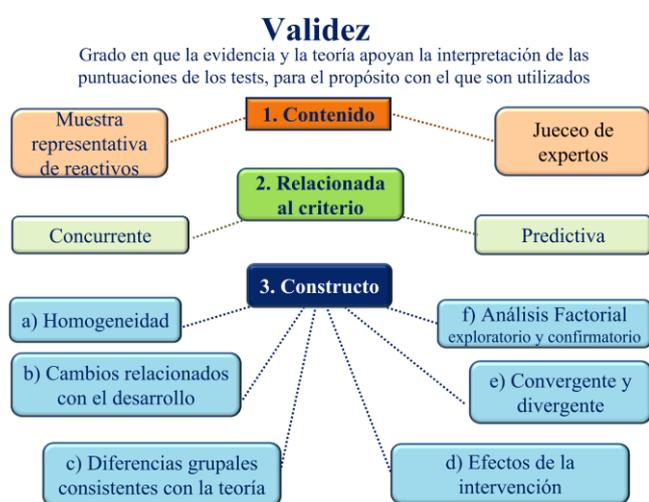


Figura. 1. Tipos de validez y técnicas o procedimientos para determinarla

La validez relacionada con el constructo como se observa en la Figura 1, es la que se puede demostrar con un mayor número procedimientos, si bien el análisis factorial exploratorio (AFE) o confirmatorio (AFC) son los más usados existen otros que contribuyen a demostrar esta propiedad psicométrica valiosa. A continuación, se describen cinco procedimientos que pueden utilizarse para analizar la validez relacionada con el constructo. El procedimiento conocido como homogeneidad de la prueba, debe demostrar que los indicadores son uniformes para evaluar el constructo, entonces se calculan las correlaciones entre cuatro diadas: ítem-ítem, ítem-escala, escala-escala, escala-total, en todos los casos se esperan asociaciones con magnitudes significativas desde moderadas a elevadas de acuerdo con la aportación de los indicadores, subescalas y puntuaciones totales al validar el constructo. [3,6]

Finalmente, el AFE o AFC demuestran la estructura factorial de la prueba, ya sea unidimensional (unifactorial) o multidimensional, considerando la varianza y el error para explicar el constructo a validar con la prueba. Ambas técnicas deben ser complementarias, así el AFE precede al AFC, en virtud de que en el primero se exploran y comparan los factores del constructo a evaluar con una teoría, mientras que, en el segundo se confirma la estructura derivada del AFE. [8,9,16]

Será complejo demostrar que una prueba posee todas las evidencias de validez de constructo, Gregory [6] ha referido que es mejor demostrar una técnica sólida en lugar de demostrar varias evidencias débiles, por ello es que, tanto el AFE como el AFC son técnicas robustas, [14, 15, 16] Sin embargo, se puede dar el caso de pruebas valiosas en el campo de la psicología en las que su estructura factorial no es estable transculturalmente, lo cual indica que el constructo puede influir factores culturales y confirma la importancia de analizar la validez

de un instrumento en las muestras en las que se aplica. Cuando eso ocurre se sugiere apoyarse en el análisis de otros procedimientos psicométricos.

Una vez que se ha generado evidencia respecto a la validez de un instrumento es imperante demostrar que también es confiable. [14, 15]

### Confiabilidad

En cuanto al atributo de confiabilidad entendido como la consistencia con la que los indicadores de una prueba producen resultados semejantes independientemente de quién o cómo se aplique. Existen dos formas para determinarla, una es de consistencia interna y otra de estabilidad temporal. En la Figura 2, se observa que en la primera forma se puede determinar con el coeficiente *Alfa de Cronbach*, el de *Kuder-Richardson* o el *Omega de Mc Donald* y, otra técnica es la de mitades (*Split half*), [15] entendiéndose que un valor superior al .70 es sinónimo de confiabilidad aceptable, en tanto que, un valor superior a .90 es confiabilidad excelente.

El coeficiente *Alfa de Cronbach* es el más utilizado. No obstante, se han dirigido críticas [17] al uso de este coeficiente ya que, puede sesgar los datos al usarse en variables de tipo ordinal, que además dependen del número de reactivos y opciones de respuesta, por ello actualmente, la confiabilidad también puede analizarse con el coeficiente *Omega de Mc Donald*, que solventa las limitaciones observadas en el *Alfa de Cronbach*. [18]

la prueba, los datos derivados se deben dividir en dos partes iguales. Los paquetes estadísticos derivan las dos mitades seleccionando aleatoriamente los indicadores que conforman cada mitad. [6] Posteriormente, se debe aplicar un coeficiente de correlación para determinar el grado de asociación entre dichas mitades, no obstante, es importante ajustar dicho cálculo con otro coeficiente, como es el de *Spearman-Brown* para obtener el resultado final de la asociación de las mitades. [6]

En algunas pruebas consideradas como “subjetivas, se requiere demostrar la confiabilidad con el dictamen de dos o más jueces. Por lo que se calcula un coeficiente de acuerdo entre calificadores para orientarse a la objetividad. [6,9]

La otra forma para determinar la confiabilidad es la de estabilidad temporal, aquí el tiempo es un aliado para demostrar que las puntuaciones de un instrumento son estables a través del tiempo, a diferencia de las técnicas para demostrar la consistencia interna (Ver Figura 2). En este procedimiento conocido como *Test-Retest* se requiere llevar a cabo dos aplicaciones de la misma prueba considerando un tiempo que no afecte o genere un sesgo en la medida. Este tipo de confiabilidad se demuestra aplicando un coeficiente de correlación a los datos derivados de los dos momentos de aplicación, es obvio que debe esperarse un coeficiente de magnitud elevada ya que, se trata de la misma prueba. [19]

La técnica de formas paralelas implica un procedimiento más complejo porque una vez que se ha seguido todo el largo camino científico del diseño de una prueba, se debe crear una forma alterna que evalúe exactamente lo mismo. Una vez que se tienen las dos formas paralelas se aplica un coeficiente de correlación esperando encontrar un coeficiente de magnitud elevada, por tratarse de la misma prueba. [6]

Como se ha referido, una prueba debe demostrar conjuntamente su objetividad, validez y confiabilidad para ser considerada como un instrumento de evaluación científicamente útil (Ver Figura 3). No obstante, el siguiente paso es emplear de manera ética y responsable la prueba, ya que, a partir de dichos datos se tomarán decisiones importantes, siendo este último paso el que contribuye a la psicología en el ámbito aplicado. Así, cuando se redacta un informe de investigación se debe explicitar nombre del instrumento, autores, año y sus propiedades psicométricas tanto la confiabilidad como la validez.

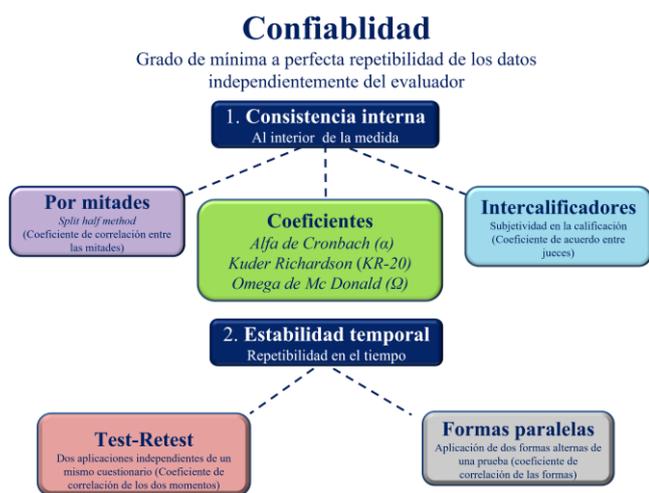


Figura 2. Tipos de confiabilidad y procedimientos para determinarla

Concretamente la técnica de división por mitades es un auxiliar o complemento de las otras técnicas de confiabilidad. Para calcularla, una vez que se ha aplicado

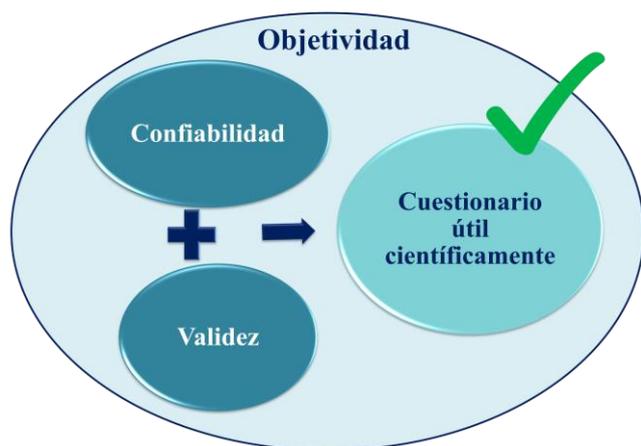


Figura 3. Objetividad, confiabilidad y validez, atributos necesarios para que un cuestionario sea útil científicamente

## Conclusiones

Los instrumentos psicométricos son una herramienta valiosa para los profesionales de la psicología y de otras ciencias sociales en virtud de su practicidad para obtener cantidades considerables de datos de manera rápida. No obstante, es una práctica ética demostrar que los instrumentos que se usarán para hacer evaluaciones en cualquier ámbito, ya sea clínico, salud, organizacional, investigación, educación regular y de necesidades especiales deben poseer las garantías científicas de la objetividad, la validez y la confiabilidad. Para analizar la validez y la confiabilidad existen un número importante de procedimientos para demostrar dichos atributos, como se ha señalado líneas arriba. Por ello, es importante poseer conocimientos de la teoría psicométrica y de estadística.

La ciencia en general experimenta un avance vertiginoso en cuanto a tecnología, no obstante, a la par se viven situaciones históricas de salud como la que ha generado la pandemia por la COVID-19, llevando a la población mundial y nacional a desarrollar alteraciones psicológicas que deben ser tratadas, por lo cual el primer paso es la evaluación oportuna, que se logra a través de instrumentos psicométricos objetivos, válidos y confiables, sin perder de vista que para elaborar un diagnóstico psicológico preciso, se debe contar con diversos medios de evaluación.

## Referencias

[1] Hernández R, Mendoza C. Metodología de la Investigación. Las rutas cuantitativa, cualitativa y mixta. México: Mc Graw Hill Educación; 2018.

[2] Tristán A, Pedraza NY. La objetividad en las pruebas estandarizadas. Rev. Iberoam. Eval. Educ. [Internet]. 2017 [Consultado 1 Nov 2022]; 10(1):11- 31. Disponible en: <https://doi.org/10.15366/riee2017.10.1.001>

[3] Hogan T. Pruebas psicológicas: una introducción práctica. (2ª ed.). México: Manual Moderno; 2015.

[4] Arrimada M. Teoría clásica de los tests: qué es y qué explica. Psic. Mente. [Internet] 2022 [Consultado 1 Nov 2022]; Disponible en: <https://psicologiymente.com/psicologia/teoria-clasica-test>

[5] Muñiz J. Las teorías de los tests: Teoría clásica y Teoría de respuesta a los ítems. Papeles Psic. [Internet] 2010[Consultado 1 Nov 2022]; 31(1); 57.61. Disponible en: <https://www.papelesdelpsicologo.es/pdf/1796.pdf>

[6] Gregory R. Evaluación psicológica: Historia, principios y aplicaciones. México: Manual Moderno; 2001.

[7] Fernández-Ballesteros R. Conceptos y modelos básicos. En Fernández-Ballesteros R, editor. Evaluación psicológica: conceptos, método y estudios de caso Madrid: Pirámide; 2013. p. 27-59.

[8] Caplan RM, Zaccuzzo DP. Pruebas psicológicas. Principios, aplicaciones y temas. 6th. ed. México: Thomson; 2006.

[9] Bautista-Díaz ML, Franco-Paredes K, Díaz-Reséndiz FJ, Ortega-Andrade NA. Los cuestionarios como un medio para realizar evaluación docente: elementos de psicometría. En Hickman, H y Alarcón, ME (Coords.), La evaluación en el espacio universitario: apuntes. México: Universidad Nacional Autónoma de México; 2019, p. 33-60.

[10] Escobar J, Cuervo A. Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances Med.* [Internet] 2008 [Consultado 1 Nov 2022]; 6(1): 27-36. Disponible en: [http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3\\_Juicio\\_de\\_expertos\\_27-36.pdf](http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf).

[11] Galicia LA, Balderrama JA, Edel R. Validez de contenido por juicio de expertos: propuesta de una herramienta virtual. *Apertura.* [Internet] 2017 [Consultado 1 Nov 2022]; 9(2): 42-53. Disponible en: <http://dx.doi.org/10.32870/Ap.v9n2.993>.

[12] Pedrosa I, Suárez-Álvarez J, García-Cueto, E. Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Accion Psic.* [Internet] 2010 [Consultado 1 Nov 2022]; 10(2): 3-20. Disponible en: [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1578-908X2013000200002](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1578-908X2013000200002)

[13] Sireci S, Faulker M. Validity evidence based on test content. *Psicothema.* [Internet] 2014[Consultado 1 Nov 2022]; 26(1): 101-107. Disponible en: <https://www.psicothema.com/pdf/4167.pdf>

[14] Díaz PA, Leyva E. Metodología para determinar la calidad de los instrumentos de evaluación. *Educ. Med. Super.* [Internet] 2013 [Consultado 1 nov 2022]; 27(2):269-86. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412013000200014&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412013000200014&nrm=iso)

[15] Ventura, JL. La importancia de reportar la validez y confiabilidad en los instrumentos de medición: Comentarios a Arancibia et al. *Rev. méd. Chile.* [Internet] 2017 [Consultado 1 Nov 2022]; 145 (7). Disponible en: [https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0034-98872017000700955](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872017000700955)

[16] Pérez E., Medrano L. Análisis factorial exploratorio: bases conceptuales y metodológicas. *Rev. Arg. Comp.* [Internet]

2010[Consultado 1 Nov 2022]; 2(1): 58-66. Disponible en: <https://revistas.unc.edu.ar/index.php/racc/article/view/15924>

- [17] Oliden PE, Zumbo BD. Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*. [Internet] 2008 [Consultado 1 Nov 2022] 20(4); 896-901. Disponible en: <https://www.redalyc.org/pdf/727/72720458.pdf>
- [18] Ventura-León JL, Caycho-Rodríguez T. El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. *Rev. Latinoam. Cien. Soc. Niñ. Juv.* [Internet] 2017 [Consultado 1 Nov 2022]; 15(1): 625-627. Disponible en: <https://www.redalyc.org/pdf/773/77349627039.pdf>
- [19] Reidl-Martínez LM. Confiabilidad en la medición. *Inv. Educ. Med.* 2013. [Internet] 2(6) [Consultado 1 Nov 2022]; 107-111. Disponible en: <https://www.redalyc.org/pdf/3497/349733227007.pdf>