

Documento de reflexión no derivado de investigación

# Calidad psicométrica de un instrumento: breve revisión

## *Psychometric quality of an instrument: brief review*

Recibido: 5 de octubre de 2022 / Aceptado: 20 de marzo de 2023 / Publicado: 1 de septiembre de 2023

Nathaly Berrío García\*, Karina Paola Zedán-Salinas\*\*

**Forma de citar este artículo en APA:**Berrío García, N., & Zedán-Salinas, K. P. (2023). Calidad psicométrica de un instrumento: breve revisión. *Poiésis*, (45), 95-103. <https://doi.org/10.21501/16920945.4522>

### Resumen

El objetivo del presente escrito es revisar las características técnicas y las propiedades psicométricas que deben tener los instrumentos psicológicos, para su uso seguro y ético. Las propiedades psicométricas de los ítems de una prueba incluyen seis aspectos que deben ser medidos adecuadamente y en su totalidad para que la evaluación de calidad técnica sea completa: dificultad y discriminación de los ítems, evidencias de validez y confiabilidad, homogeneidad de los ítems, análisis de distractores y funcionamiento diferencial de los ítems. Se concluye que la calidad técnica de una prueba debe estar relacionada con la elección de la misma por parte de los psicólogos, y que es una decisión ética de quienes tienen la responsabilidad de realizar un acompañamiento objetivo y consistente basado en la evidencia, independientemente del campo laboral en que se desempeñen.

### Palabras clave:

Confiabilidad; Instrumentos; Medición; Método de evaluación; Psicometría; Test psicológico; Validez.

\* Doctora en Epidemiología coordinadora del Grupo Problemas Clínicos y Psicosociales en la Corporación Universitaria Empresarial Alexander von Humboldt, Armenia, Colombia. Contacto: [nberrio18@cue.edu.co](mailto:nberrio18@cue.edu.co) ORCID: <https://orcid.org/0000-0002-3574-4101>

\*\* Especialista en Gerencia de Proyectos, pasante de investigación en el Programa +Mujer +Ciencia +Equidad, Armenia, Colombia. Contacto: [ka.pa.ze.sa@gmail.com](mailto:ka.pa.ze.sa@gmail.com) ORCID: <https://orcid.org/0000-0002-5469-5560>

## Abstract

The objective of this paper is to review the technical characteristics and psychometric properties that psychological instruments must have, for their safe and ethical use. The psychometric properties of test items include six aspects that must be adequately and fully measured for the evaluation of technical quality to be complete: item difficulty and discrimination, validity and reliability evidence, item homogeneity, analysis of distractors and differential functioning of the items. It is concluded that the technical quality of a test must be related to the choice of the test by psychologists, and that it is an ethical decision of those who have the responsibility to carry out an objective and consistent follow-up based on evidence, regardless of the job in which they work.

## Keywords:

Reliability; Instruments; Measurement; Evaluation method; Psychometry; Psychological test; Validity.

# Introducción

El uso de instrumentos de evaluación y medición en psicología es cada vez más frecuente. A la hora de elegir cualquier tipo de herramienta es importante considerar tanto la pertinencia como la calidad de esta, es decir, asegurarse que la medición sea lo más cercana posible a la puntuación real que se intenta medir y que dicha medición sea consistente en el tiempo. Es por esto que la construcción de una prueba debe cumplir con criterios rigurosos de calidad, que sean de utilidad. La calidad referida es producto de cada componente de una escala o prueba, la cual tiene características que dan cuenta de las evidencias de confiabilidad y validez con las que fue diseñada y cuyo objetivo es cuantificar o cualificar la o las variables a las que se apunta (Cohen & Swerdlik, 2001; García-Portilla et al., 2022).

De acuerdo con los valores encontrados en la ficha técnica o a lo largo del manual de la prueba, se puede juzgar sobre las propiedades psicométricas generales o por dominio y, a partir de allí, tomar decisiones objetivas en el ejercicio profesional. Es responsabilidad de los psicólogos conocer detalladamente la información contenida en los documentos mencionados y con ello decidir usar pruebas psicométricas afines con los objetivos de evaluación; así como encontrarse capacitado para la aplicación e interpretación adecuada de las puntuaciones de la prueba, conocimientos estadísticos que subyacen la construcción del instrumento y los modelos teóricos que la soportan; por último, el respeto a los derechos del evaluado y seguir reglas o principios éticos nacionales o internacionales asociados con la evaluación psicológica (Ocampo, 2003; Consejo General de la Psicología de España, 2022).

Por tanto, para cumplir eficientemente con esta labor, los profesionales de la psicología deben tener presentes los parámetros recomendables para la construcción o elección de una prueba psicométrica de calidad. De allí que el objetivo de esta revisión es revisar las características técnicas y las propiedades psicométricas que deben tener los instrumentos psicológicos para su uso seguro y ético.

## *Calidad psicométrica o técnica*

La *calidad psicométrica* o *calidad técnica* de un instrumento es un atributo que abarca tanto las evidencias de confiabilidad y validez de una prueba, como los criterios de baremación y los procesos con los que se diseñó dicho instrumento (Nuván Hurtado et al., 2018). Por tanto, la evaluación de la calidad psicométrica de los elementos de la prueba permite identificar las características estadísticas de estos y su aporte a la medición de las variables para las que fue diseñada.

En consecuencia, las propiedades psicométricas de los ítems de una prueba incluyen seis aspectos que deben ser medidos adecuadamente y en su totalidad para que la evaluación de calidad técnica sea completa:

- Dificultad de los ítems.
- Discriminación de los ítems.
- Evidencias de validez y confiabilidad.
- Homogeneidad de los ítems.
- Análisis de distractores.
- Funcionamiento diferencial de los ítems (American Educational Research Association et al., 2014; Romero & Ordoñez, 2015).

### *Dificultad de los ítems*

El *índice de dificultad* de los ítems representa la proporción de evaluados que lo responden de forma correcta. Como se refiere a una proporción, puesto que los acertantes son una parte de los respondientes, sus valores oscilan entre 0 y 1, y con frecuencia se expresan con un porcentaje. Los índices cercanos a 1 sugieren una baja dificultad, e índices próximos a 0 señalan una dificultad máxima. El índice debe ser corregido para contemplar que una parte de los aciertos puede deberse al azar. La corrección debe incluir en el cálculo el número de alternativas de respuesta. Adicionalmente, la dificultad del ítem debe estar relacionada con la dificultad del concepto que está evaluando. Este aspecto es fundamental en la calidad técnica, porque si la mayoría de los respondientes fallan en el ítem, es porque su índice de dificultad no es adecuado y se les está penalizando de manera inadecuada (Meneses et al., 2003; Romero & Ordoñez, 2015).

### *Discriminación de los ítems*

Es la capacidad que ofrecen los ítems de distinguir entre los evaluados que tienen un buen rendimiento en la prueba respecto a los que tienen uno malo. Si los ítems no discriminan, no son útiles para la medición, y ese es el propósito para el que fueron redactados e incluidos en la prueba. Es decir, un ítem debe ser más acertado entre las personas que han obtenido una alta puntuación en el instrumento que entre los que tienen una baja puntuación. También puede ser expresado en porcentaje. Los métodos para establecer quiénes conforman los grupos de mejor y peor rendimiento son variados e incluyen el uso de la media, la mediana, los cuartiles, entre otros. Este método influirá en la toma de decisión respecto a cuáles ítems deben ser conservados o retirados del instrumento. En general, los ítems con índices de discriminación inferiores al 20% son inaceptables y deben ser eliminados del instrumento. Asimismo, es importante resaltar que el índice de discriminación

depende estadísticamente de (la varianza de) la dificultad y no debe interpretarse *per se*. Es decir, que es altamente probable que un ítem que no discrimine bien es porque su dificultad es inadecuada para la medición de la variable (Meneses et al., 2003; Romero & Ordoñez, 2015).

## *Evidencias de validez y confiabilidad*

La validez evalúa que la prueba mida la variable que pretende medir. Tiene que ver con qué tan bien mide la variable el instrumento y qué se puede inferir a partir de sus resultados (Anastasi & Urbina, 1998; Hogan, 2015). Son múltiples las evidencias de validez que se evalúan en una prueba:

- Evidencia basada en el contenido de la prueba.
- Evidencia basada en los procesos de respuesta.
- Evidencia basada en la estructura interna.
- Evidencia basada en relaciones con otras variables.
- Evidencia de validación y consecuencias de las pruebas (American Educational Research Association et al., 2014).

Por lo anterior, la evaluación integral de las evidencias de validez de un instrumento conlleva a que se deben medir diferentes evidencias y no solamente calcular coeficientes de validez de contenido a partir de juicios de expertos, como tradicionalmente solía realizarse (Meneses et al., 2003; Romero & Ordoñez, 2015).

Por su parte, la propiedad psicométrica de confiabilidad se refiere al nivel real del sujeto en relación con la variable medida. Una prueba confiable es aquella cuyas puntuaciones son consistentes (Anastasi & Urbina, 1998; Hogan, 2015). También son múltiples las evidencias de confiabilidad que se evalúan en una prueba:

- Evidencia de confiabilidad test-retest.
- Evidencia de confiabilidad interjueces.
- Evidencia de confiabilidad de formas alternas.
- Evidencia de confiabilidad de consistencia interna (Hogan, 2015).

De ahí que la evaluación integral de las evidencias de confiabilidad de un instrumento implica que se deben medir diferentes evidencias y no solamente calcular coeficientes de consistencia interna con Alfa de Cronbach, como tradicionalmente solía realizarse (Meneses et al., 2003; Romero & Ordoñez, 2015). Por otro lado, se debe enfatizar que la consistencia interna se mide por diversos métodos, pero que la selección del adecuado depende de la naturaleza y nivel de medición de las variables medidas, y del tipo de ítem. Por ejemplo, para ítems ordinales, es decir, con escala tipo Likert, debe emplearse coeficiente Omega o alfa ordinal, teniendo en cuenta que los análisis estadísticos se basan en una matriz de correlaciones diferente a la de Pearson: la matriz

de correlaciones policóricas. Este es un análisis que no puede realizarse mediante paquetes por defecto como el Statistical Package for the Social Sciences (SPSS), sino que deben emplearse otro tipo de softwares como el R o Jamovi (Lloret-Segura et al., 2014; Viladrich et al., 2017).

Finalmente, las propiedades psicométricas deben ser medidas de forma rigurosa antes de que el instrumento sea aplicado masivamente a sus destinatarios, por lo que los modelos de análisis deben ser evaluados mediante índices de bondad de ajuste absoluto, incremental y de la parsimonia (Escobedo Portillo et al., 2016; Salas Vargas et al., 2017).

## *Homogeneidad de los ítems*

Se trata de la medición del grado de acuerdo entre los ítems para verificar si miden la misma variable. Si el instrumento tiene subescalas o subtest, es decir, se compone de varios apartados con reactivos o ítems que miden variables distintas, se debe evaluar la homogeneidad de los ítems de cada subescala y respecto a la prueba completa (Gómez Mejías et al., 2018).

## *Análisis de distractores*

*Distractores* es el nombre técnico para las alternativas de respuesta incorrectas en un reactivo. Un aspecto cardinal para el adecuado funcionamiento de un ítem es que sus distractores realmente lo sean (Meneses et al., 2003; Romero & Ordoñez, 2015). Un buen distractor no debe ser elegido por más del 10% de los evaluados, por lo que su promedio debe ser menor que el promedio de la opción de respuesta correcta (Romero & Ordoñez, 2015). Igualmente, se debe calcular el índice de discriminación para cada distractor. “Si un ítem tiene una opción de respuesta inverosímil (por ejemplo, Maradona como autor de El Quijote), el ítem será más fácil y necesariamente discriminará peor” (Meneses et al., 2003, p. 246).

## *Funcionamiento diferencial de los ítems*

El objetivo de este análisis es identificar ítems cuya probabilidad de acierto sea diferente entre subgrupos de evaluados, que tienen igual desempeño en las variables medidas en la prueba, es decir, identificar y eliminar sesgos que puedan existir contra grupos con determinadas características sociodemográficas o de otra índole. Existe funcionamiento diferencial de un ítem cuando la probabilidad de responder correctamente no depende solamente del nivel de habilidad de los evaluados, sino de su pertenencia a diferentes grupos. Esto es, un ítem tiene funcionamiento diferencial cuando dos grupos de evaluados que presentan idéntico nivel de la variable medida por la prueba obtienen resultados diferentes en la ejecución de la misma (Meneses et al., 2003; Romero & Ordoñez, 2015).

Por último, debe tenerse en cuenta que una inadecuada evaluación de la calidad técnica de un instrumento conduce a sesgos en la medición, con la consecuente toma de decisiones errada a partir de los resultados obtenidos por los evaluados.

## Conclusiones

La calidad técnica de una prueba parte de modelos matemáticos y análisis estadísticos que describen parámetros de confiabilidad y validez para definir la pertinencia y utilidad de un instrumento. Dicha medición se puede establecer para la prueba en general y por dominios o dimensiones de una variable. Por tal razón, la elección de los ítems que componen una prueba no debe ser azarosa, más bien es importante recurrir a indicadores de evaluación como los seis aspectos referidos anteriormente para predecir la calidad psicométrica del instrumento, según el grado de acierto y error que se identifique en la construcción de la misma.

Recuérdese que también permite predecir si la medición corresponde a la discriminación entre las características generales y excepcionales de un atributo, por lo que su uso en diversas áreas de la psicología permite optimizar y dar veracidad a la labor de los profesionales. La elección de las pruebas por parte de los psicólogos es una decisión ética de quienes tienen la responsabilidad de realizar un acompañamiento objetivo y consistente basado en la evidencia, independientemente del campo laboral en que se desempeñen.

## Conflicto de intereses

Las autoras declaran la inexistencia de conflicto de interés con institución o asociación comercial de cualquier índole.

# Referencias

- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *Estándares para Pruebas Educativas y Psicológicas*. American Educational Research Association. [https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302745\\_web.pdf](https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302745_web.pdf)
- Anastasi A. y Urbina, S. (1998). *Tests psicológicos*. Pearson Educación.
- Cohen R. y Swerdlik, M. (2001). *Pruebas y evaluación psicológica: introducción a las pruebas y a la medición* (J. A. Vásquez Arellano, Trad.). Editorial McGraw Hill.
- Consejo General de la Psicología de España. (s.f.). *Comisión de Test*. <https://www.cop.es/index.php?page=principios-eticos>
- Escobedo Portillo, M. T., Hernández Gómez, J. A., Estebané Ortega, V. y Martínez Moreno, G. (2016). Modelos de ecuaciones estructurales: Características, fases, construcción, aplicación y resultados. *Ciencia & Trabajo*, 18(55), 16-22. <https://doi.org/10.4067/S0718-24492016000100004>
- García-Portilla, M., Bascarán, M., Sáiz, P., Bobes, M., Bousoño, M., & Bobes, J. (2022). *Banco de instrumentos básicos para la práctica de la psiquiatría clínica*. Editorial Universidad de Oviedo.
- Gómez Mejías, A. E., Cerrada Valero, R., & Rangel Vasquez, R. E. (2018). Validez del material educativo de un programa de educación ambiental- sanitario no formal. *Educere*, 22(71), 131-152. <http://erevistas.saber.ula.ve/index.php/educere/article/view/12085/21921923196>
- Hogan, T. P. (2015). *Pruebas psicológicas: una introducción práctica*. El Manual Moderno.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El Análisis Factorial Exploratorio de los Ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L.M., Turbany, J., & Valero, S. (2003). *Psicometría*. Editorial UOC.
- Nuván Hurtado, I. L., Rivera Porras, D. A., Carrillo Sierra, S. M., Forgiony Santos, J. O., Bonilla Cruz, N. J. y Roza Sánchez, A. C. (2018). Diferencias en la calidad psicométrica de test construidos mediante la estrategia pedagógica audiovisual y las estrategias pedagógicas tradicionales. *Revista Espacios*, 39(25). <https://bonga.unisimon.edu.co/handle/20.500.12442/2310>
- Ocampo, L. (2003). Responsabilidad ética y profesional frente al uso de las pruebas psicométricas como herramientas de evaluación psicológica. *Revista Universidad de San Buenaventura Medellín*, 18, 147-149.

Romero, S. J., & Ordoñez, X. G. (2015). *Psicometría*. Ediciones Centro de Estudios Financieros Universidad a Distancia de Madrid.

Salas Vargas, A. A., Navarro Camacho, R., & Montero Rojas, E. (2017). Un modelo de ecuaciones estructurales para el estudio de factores que afectan la competencia lectora y la alfabetización matemática: Una aproximación bayesiana con datos de PISA 2009. *Estadística Española*, 59(194), 167-192. <https://www.kerwa.ucr.ac.cr/handle/10669/82695>

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33(3), 755-782. <https://doi.org/10.6018/analesps.33.3.268401>