

Observaciones para la construcción y validación de instrumentos de investigación

Observations for Construction and Validation of Research

Herman Tarazona Mirabal ^{1, a}
<http://orcid.org/0000-0001-5319-4708>

Recibido: 01-09-2020

Aceptado: 30-10-2020

Citar como

Tarazona Mirabal, H. (2020). Observaciones para la construcción y validación de instrumentos de investigación. *Desafíos*, 11(2), 177-82. <https://doi.org/10.37711/desafios.2020.11.2.213>

RESUMEN

El diseño de un instrumento de medición es una tarea multidisciplinaria, aún para medir algún rasgo sin mayores implicancias; dependiendo de la aplicación que se le dé va a necesitar especialistas en el constructo, analistas estadísticos, diseñadores de test, especialistas en el lenguaje. De hecho, se requiere un conocimiento especializado para el diseño o adecuación de instrumentos que tengan amplias repercusiones como, por ejemplo, validar un instrumento de diagnóstico médico. Las características métricas de un instrumento son su validez y confiabilidad. La validez hace referencia a la exactitud de instrumentos físicos y la confiabilidad a la precisión en la medida. Las valideces que son representativas de esta característica son la validez de contenido, de constructo y de criterio, mientras que la confiabilidad del instrumento aumenta cuando están identificados los errores y estos se pueden controlar.

Palabras clave: instrumento, constructos, validez, confiabilidad.

ABSTRACT

Design of a measurement instrument is a multidisciplinary task. To measure some trait without major implications - depending on the application you are given- it will be needed specialists in construct, statistical analysts, test designers, and language specialists. In fact, specialized knowledge is required for the design or adaptation of instruments that have broad repercussions, such as, validating a medical diagnostic instrument. Metric characteristics of an instrument are its validity and reliability. Validity refers to the accuracy of physical instruments and the reliability of measurement precision. In addition, validities that are representative of this characteristic are content, construct and criterion validity, while reliability of the instrument increases when errors are identified and these can be controlled.

Keywords: instrument, constructs, validity, reliability.

Filiación y grado académico

¹ Universidad de Huánuco, Huánuco, Perú.

^a Magister en Educación.



INTRODUCCIÓN

La investigación, en cualquier área del conocimiento, trata de observar o medir el comportamiento de las variables. En el caso de las ciencias de la naturaleza las variables generalmente están identificadas y éstas se pueden observar y medir, aunque en algunos casos su observación no es directa y se miden sus manifestaciones.

En el caso de las variables o conceptos del comportamiento, generalmente no contamos con instrumentos físicos o equipos que midan o detecten el fenómeno que se está observando, pues hasta para precisar el constructo debe definirse en base a estudios teóricos, opiniones de expertos y antecedentes de investigación (Del Carpio, 2019).

El concepto de un comportamiento puede recibir diversos sinónimos, como constructo, variable conceptual, concepto teórico; y su definición está vinculada a una teoría (Reguant, 2014). Son variables de este tipo la inteligencia, la motivación, el aprendizaje, la personalidad, el éxito escolar, la disfunción eréctil, la ansiedad, etc. Las variables también son propiedades que asumen diversos valores (Kerlinger, 1985); de este modo, algunas variables pueden adoptar más de dos valores, como actitud hacia la política (alto, moderado, indiferente, despectivo), o pueden tomar solo dos valores, como el caso de infección (positivo, negativo). Existen variables que adoptan infinitos valores, como la cantidad de glucosa en la sangre.

Para medir los conceptos o constructos teóricos estos deben ser operacionalizados; lo que se entiende como las actividades necesarias para la ejecución de una tarea o el registro de manifestaciones observables frente al fenómeno de interés. La operacionalización de variables o conceptos es un procedimiento de partición de los aspectos abstractos del concepto, hasta llegar a niveles concretos de la realidad (Reguant, 2014). Por ejemplo, si se estudia el concepto "actitudes hacia el uso de la tecnología para el aprendizaje de las matemáticas", se determinan cuatro dimensiones o factores: utilidad de la tecnología en la matemática, actitud hacia el aprendizaje de matemáticas con la tecnología, experiencia de matemáticas con la tecnología y matemáticas y tecnología (López, 2007).

De allí la necesidad de construir instrumentos de medición de los conceptos, que son dispositivos estructurados que deben desagregar el concepto en manifestaciones observables a las cuales se les asignará un puntaje. El instrumento de investigación

es una herramienta para recoger los datos observables de una variable, ya sea por medio de un test, un cuestionario, una entrevista (Soriano, 2014).

Estado del arte

En el proceso de diseño del instrumento se asigna puntajes a las manifestaciones de la variable, cuya elección se fundamenta en bases teóricas, epistemológicas y paradigmas. Existen diversas propuestas de investigadores para la construcción de un instrumento; por ejemplo, para el diseño de un test en el dominio afectivo se deben considerar las siguientes etapas: 1) identificación de la finalidad del test, 2) establecer el marco teórico del constructo, 3) especificaciones derivadas del marco, 4) revisión del constructo y de los ítems, 5) estudio piloto, 6) análisis de la información y diseño final del instrumento (Martínez, et al., 2014, p. 25).

Soriano (2014, p.19), por su parte, propone: 1) redacción de ítems, 2) juicio de expertos, 3) prueba piloto, 4) validación psicométrica. De igual modo, Carvajal et al., (2011) proponen para el desarrollo de instrumentos de la salud las siguientes etapas: 1) Definir el objetivo, 2) estudio de la literatura, 3) revisión de instrumentos existentes, 4) definición del constructo, 5) especificación de las dimensiones, 6) definiciones operativas y redacción de ítems, 7) definiciones conceptuales, 8) revisión por expertos, 9) dirección de respuestas, piloto inicial, 10) análisis de datos, 11) segundo piloto, 12) generación del instrumento final, 13) determinación de fiabilidad y validez.

Finalmente, Supo (2013, p. 7) establece diez pasos para construir y validar un instrumento: 1) revisión de la literatura y definición del concepto, 2) entrevista no estructurada, 3) listar los temas, 4) construcción de reactivos, 5) selección de jueces, 6) pasar la prueba piloto, 7) determinar la validez interna y la de criterio, 8) reducir ítems, 9) definir las dimensiones, 10) comparar con otro criterio.

Validación de instrumentos

La validación de un instrumento comprende: 1) La validación lógica del constructo, a cargo de expertos; lo que implica la validación de la definición teórica del constructo, la validación de la definición operativa, el diseño y validación de los ítems. 2) La validación métrica del instrumento mediante un estudio piloto; lo que implica el análisis estadístico descriptivo, de discriminación, de dimensionalidad y de calidad de los reactivos (Bakieva, 2016).

Asimismo, la calidad de un instrumento comprende tanto su validez como su confiabilidad. La validez se refiere a si el instrumento mide lo que sus autores establecen (Covacevich, 2014).

Manteniendo los principales componentes de la validez y con el objetivo de que la descripción pueda ser una guía para la validación de instrumentos, se consideran los siguientes tipos de validez: 1) validez de contenido, 2) validez de constructo, 3) validez de criterio, 4) validez discriminante y 5) validez divergente. (Zegarra, 2011). Enfatizamos ahora los tres primeros tipos:

Validez de contenido

Para establecer los ítems, “se requiere tener una adecuada conceptualización y operacionalización del constructo, es decir, el investigador debe especificar previamente las dimensiones a medir y sus indicadores, a partir de los cuales se realizarán los reactivos” (Escobar y Cuervo, 2008, p.2)

Los expertos o jueces realizan la validez de contenido. El grado de acuerdo entre los jueces se estima mediante el coeficiente de concordancia W de Kendall, para datos en escala ordinal. Para datos nominales se utiliza el estadístico Kappa.

Escobar y Pérez (2005) sugieren las siguientes tablas de evaluación de los ítems (a continuación) por parte de los jueces; lo que nos permite procesar el grado de acuerdo estableciendo tres criterios para la evaluación de los reactivos:1) La claridad de un ítem significa que éste es comprensible en sintáctica y semántica, 2. La coherencia indica la relación lógica con la dimensión que se está evaluando, 3. La relevancia indica la importancia del ítem. Para la evaluación de las dimensiones se utiliza el criterio de suficiencia; lo que indica que los reactivos de una misma dimensión la evalúan adecuadamente. Se cita como ejemplo las siguientes tablas para la validez de contenido:

Validez del constructo

Se puede considerar como validez basada en la estructura interna, como los ítems pueden

Tabla 1
Datos del experto y del proyecto

Plantilla para evaluar la validez del contenido a través del juicio de expertos													
Proyecto: (ejemplo) El aprendizaje cooperativo y las habilidades crítico reflexivas													
Objetivo: Demostrar la eficacia del aprendizaje cooperativo para desarrollar habilidades crítico reflexivas													
Nombre del Experto:													
Formación Académica:													
Áreas de experiencia profesional:													
Cargo Actual:													
Institución:													
Nota. Ejemplo de proyecto de R. Marroquín (2006)													

relacionarse con una o más dimensiones que se consideran como variables latentes no directamente observables. Además, para determinar la estructura interna de cualquier escala de un test y para encontrar el significado teórico a un conjunto inicial de ítems, se utiliza el análisis factorial.

El análisis factorial exploratorio (AFE) es un método multivariado que permite agrupar las variables que se correlacionan fuertemente entre sí. Mediante el AFE, la variabilidad de las puntuaciones de un conjunto de variables es explicada por un número más reducido de dimensiones o factores que expliquen la mayor cantidad de varianza común; además, permite identificar los ítems que no son significativos o que son redundantes (Pérez, 2010).

Las fases características del análisis factorial son: 1. cálculo de una matriz capaz de expresar la variabilidad conjunta de todas las variables, 2. extracción del número óptimo de factores, 3. rotación de la solución para facilitar su interpretación, 4. estimación de la puntuación con las nuevas dimensiones (Macía, 2010).

Tabla 2
Ejemplo de valoración de ítems (puntaje de 1 a 4)

id	Ítems (ejemplos)	coherencia				Relevancia				claridad			
		No cumple con el criterio	Bajo nivel	Moderado nivel	Alto nivel	No cumple con el criterio	Bajo nivel	Moderado nivel	Alto nivel	No cumple con el criterio	Bajo nivel	Moderado nivel	Alto nivel
1	Me conformo con la información que me dan sin importar de que fuente provengan												
2	No arribo a una conclusión sin antes haber revisado los pormenores de la información												

Tabla 3
Ejemplo de evaluación de dimensión de suficiencia

Evaluación de Suficiencia					Observaciones
Dimensiones	1. No cumple con el criterio	2. Bajo nivel	3. Moderado nivel	4. Alto nivel	
Habilidad crítica constructiva					
Habilidad reflexiva					

La varianza (σ^2) de cada ítem representa la variabilidad en las respuestas; cada ítem o variable tiene su varianza (diferencias en las respuestas), pero esta varianza de cada ítem puede estar compartida con la varianza de otros ítems (Morales, 2013).

La otra técnica que se utiliza pero para confirmar las dimensiones propuestas inicialmente del análisis teórico, es el análisis factorial confirmatorio (AFC), donde el investigador especifica de antemano el número de factores y la relación entre los indicadores y las cargas factoriales (Herrero, 2010).

Validez de criterio

Este criterio de validez también se conoce como validez en relación con otras variables, además de validez convergente o concurrente (Covacevich, 2014). Esta validez se determina comparando los resultados obtenidos del instrumento nuevo con otro test, o método, que mide el mismo constructo o constructos similares. Se espera que los puntajes de un cierto instrumento se correlacionen con otros que declaran medir constructos iguales o parecidos (Wilson, 2005). La determinación de la concordancia se realiza por medio del estadístico kappa (Zegarra, 2011).

Confiabilidad de un instrumento

La confiabilidad de un instrumento es la consistencia de la medida del rasgo o característica para el cual fue diseñado, teniendo en cuenta los posibles errores que pueden alterar la medición.

El manual del test debe identificar las principales fuentes de error, las estadísticas que cuantifiquen el tamaño de estos errores y el grado de generalizabilidad de los puntajes. (Joint Committee on Standards for Educational and Psychological Testing, 1999).

Se han desarrollado además varios marcos teóricos estadísticos importantes para analizar la confiabilidad. Los principales son la teoría clásica de medición, la teoría de generalizabilidad y la teoría de respuesta al ítem (TRI) (Haertel, 2006).

En la teoría clásica para medir la fiabilidad de la consistencia interna del instrumento se utiliza la correlación entre dos instantes de aplicación del instrumento, como en el denominado test-re-test; además se utiliza el coeficiente *alpha de Cronbach*, que se fundamenta en que los ítems miden el mismo constructo y que están fuertemente correlacionados entre sí.

La fiabilidad se define como la razón de $\sigma^2(T)/\sigma^2(X)$; es decir, es la proporción de la variabilidad en la puntuación del instrumento que se atribuye al rasgo que está siendo medido respecto a la variabilidad total que contienen las respuestas de los sujetos (el rasgo más el error) (Frías-Navarro, 2019). La fiabilidad es lo que se denomina la precisión, que es lo más usual cuando se miden variables de las ciencias naturales, y representa la reproducibilidad de la medición utilizando un instrumento o método de medición. En el caso de la precisión o fiabilidad de una variable del comportamiento, el instrumento que se aplica para la medición de algún rasgo utiliza una escala ordinal de categorización (escala Likert). Además, la muestra tiene que ser relativamente grande y puede registrarse una gran dispersión de las respuestas o comportamiento; para lo cual se vienen utilizando índices de confiabilidad como el alfa de Cronbach o el método omega de Mc Donald. La ecuación para estimar el alfa de Cronbach es:

$$\alpha = \frac{K}{K - 1} \left[1 - \frac{\sum S_i^2}{\sum S_T^2} \right]$$

Donde α es el símbolo del alfa de Cronbach, K el número de ítems, $\sum S_i^2$ es la suma de las varianzas de cada ítem y $\sum S_T^2$ es la varianza total.

El coeficiente omega (ω) de McDonald (1999) presenta ventajas sobre el alfa de Cronbach por cuanto no requiere la correlación de los errores, ni que los ítems presenten una alta correlación entre sí. Se utiliza tanto para escala ordinal, binaria o escala de intervalo; es un coeficiente que está reemplazando al alfa de Cronbach. La ecuación del coeficiente omega es:

$$\omega = \frac{[\sum_{i=1}^i \delta]^2}{[\sum_{i=1}^i \delta]^2 + \sum_{i=1}^i [1 - \delta_i^2]}$$

Dónde: ω es el símbolo del coeficiente omega; δ_i es la carga factorial estandarizada (Ventura, 2017). Dos aspectos importantes considerados como fortaleza del coeficiente omega son: 1) empleo de la adición de variables estandarizadas, 2) y no está en relación al número de ítems (Salazar y Cerpa, 2017). Las puntuaciones factoriales (δ) obtenidas del AFC se utilizan para el cálculo del coeficiente ω .

Con relación a los valores usuales del coeficiente alfa de Cronbach existen opiniones convergentes de diversos investigadores; por ejemplo, Kaplan y Saccuzzo (1982) consideran el valor de consistencia interna para la investigación básica entre 0,7 y 0,8 y en investigación aplicada, sobre 0,95.

El coeficiente alfa de Cronbach es un coeficiente para variable ordinal en escala y, además, se utiliza el coeficiente Kuder-Richardson para cuestionarios. Lo usual es aceptar valores mayores o iguales a 0,80 (Supo, 2013).

Consideraciones adicionales

Estandarización: los instrumentos deben ser aplicados bajo las mismas condiciones en las que se obtuvo su validez y confiabilidad. Esto significa que las instrucciones, el contexto de aplicación y los procedimientos de puntuación han sido exactamente los mismos para todos los examinados. (Covacevich, 2014).

Costos: los instrumentos que se disponen pueden ser de acceso libre, mientras que otros que se utilizan, por ejemplo, para evaluar procesos educativos, pueden representar costos importantes.

Facilidad de uso: los instrumentos deben tener instrucciones claras para su uso; por ejemplo, las hojas de respuesta complejas pueden confundir al usuario y un estudiante puede darse cuenta en medio del test que ha estado respondiendo en la sección equivocada (Timmons et al., 2005).

Tipos de instrumentos: se diseñan instrumentos de baja implicancia que generan información básica sin consecuencias para los administrados, mientras que los de alta implicancia utilizan sus resultados para diversos propósitos, como definir incentivos económicos (Cueto, 2007).

Los instrumentos referidos a normas y criterios se basan en la comparación del grupo evaluado con una referencia representativa; el puntaje se compara además de manera relativa en comparación con el puntaje del grupo de referencia. Los instrumentos referidos a criterios también se conocen como de contenido, competencia, dominio u objetivo (Cohen y Swerdlik, 2009).

La selección de un instrumento para un objetivo debidamente identificado, debe tener en consideración diversos aspectos, como los técnicos, económicos, éticos y prácticos; por lo que se requiere por parte de los investigadores que lo vayan a utilizar, experiencia tanto del objetivo de evaluación, así como de las propiedades técnicas del test. Desarrollar un instrumento para los propósitos definidos es por lo demás un esfuerzo que vale la pena pues está más cerca del conocimiento del objeto de evaluación y además se pueden realizar los ajustes necesarios.

REFERENCIAS

- Bakieva, M. (2016). *Diseño y validación de un instrumento para evaluar la colegialidad docente*. (Tesis Doctoral, Universidad de Valencia). <https://www.uv.es/gem/gemeduco/pub>
- Carvajal, A., Centeno, C., Watson, R., Martínez, M. y Sanz Rubiales, A. (2011). ¿Cómo validar un instrumento de medida de la salud? *Anales Sis San Navarra*, 34(1), 63-72. <http://scielo.isciii.es/pdf/asisna/v34n1/revision1.pdf>
- Cohen, R. y Swerdlik, M. (2009). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (7.a ed.). McGraw-Hill Higher Education.
- Covacevich, C. (2014). Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles (Nota Técnica # IDB-TN-738). Banco Interamericano de Desarrollo. <https://publications.iadb.org/es/publicacion/16899/como-seleccionar-un-instrumento-para-evaluar-aprendizajes-estudiantiles>
- Cueto, S. (2007). Las evaluaciones nacionales e internacionales de rendimiento escolar en el Perú: balance y perspectivas. Investigación, políticas y desarrollo en el Perú. En GRADE. Investigación, políticas y desarrollo en el Perú (pp. 405-455). Grupo de Análisis para el Desarrollo (GRADE). <http://www.grade.org.pe/download/pubs/InvPolitDesarr-10.pdf>
- Del Carpio Rivera, A. (2019). *Las variables en la investigación*. Academia.edu. https://www.academia.edu/36708945/Dra_Adela_Del_Carpio_Rivera_Doctor_en_Medicina
- Escobar, J. y Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en medición*, 6, 27-36. https://www.researchgate.net/profile/Jazmine_Escobar
- Fernández, A. (2015). Aplicación del análisis factorial confirmatorio a un modelo de medición del rendimiento académico en lectura. *Revista de ciencias económicas*, 33(2), 39-65. <https://dialnet.unirioja.es/servlet/articulo?codigo=5516149>
- Frías-Navarro, D. (2019). *Apuntes de consistencia interna de las puntuaciones de un instrumento de medida*. Universidad de Valencia. <https://www.uv.es/friasnav/AlfaCronbach.pdf>

- Herrero, J. (2010). El análisis factorial confirmatorio en el estudio de la estructura y estabilidad de los instrumentos de evaluación: Un ejemplo con el cuestionario de autoestima CA-14. *Intervención Psicosocial*, 19(3), 289-300. <https://www.redalyc.org/pdf/1798/179817507009.pdf>
- Kerlinger, F. (2002). *Investigación del comportamiento* (4ªed.). McGraw-Hill
- López, R., Castro, E. y Molina, M. (2007). *Elaboración y validación de un cuestionario de actitudes hacia el uso de la tecnología para el aprendizaje de las matemáticas*. ResearchGate. <https://www.researchgate.net/publication/279492507>
- Macía, F. (2010). *Validez de los tests y el análisis factorial: Nociones generales*. Logroño: Ciencia y Trabajo, (35), 276-280. <https://dialnet.unirioja.es/servlet/articulo?codigo=3218921>.
- Marroquín, R. (2013). Confiabilidad y validez de instrumentos de investigación. Universidad Nacional de Educación Enrique Guzmán y Valle. <http://www.une.edu.pe/Titulacion/2013/exposicion/SESSION-4>
- Martínez, M., Hernández, J. y Hernández, V. (2014). *Psicometría*. Alianza Editorial. https://issuu.com/maldonadocarla/docs/psicometr__a_-_alianza_editorial
- Morales, P. (2013). El análisis factorial en la construcción de test, escalas y cuestionarios. Universidad Pontificia Comillas. http://www.eio.uva.es/~valentin/ad3d/anadat/afc/comillas_AnalisisFactorial.pdf
- Pérez, E. y Medrano, L. (2010). *Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas*. Córdoba. *Revista Argentina de Ciencias del Comportamiento*, 2(1), 58-56 https://www.researchgate.net/publication/42091816_Analisis_factorial_exploratorio_Bases_conceptuales_y_metodologicas.
- Reguant, M. y Martínez-Olmo, F. (2014). *Operacionalización de conceptos/variables*. Diposit digital de la UB. <http://diposit.ub.edu/dspace/bitstream/2445/57883/1/Indicadores-Repositorio.pdf>.
- Salazar, C. y Serpa, A. (2017). Análisis confirmatorio y coeficiente Omega como propiedades psicométricas del instrumento Clima Laboral de Sonia Palma. *Revista De Investigación En Psicología*, 20(2), 377-388. <http://dx.doi.org/10.15381/rinvp.v20i2.14047>
- Soriano, A. (2014). *Diseño y validación de instrumentos de medición*. *Diálogos*, (14), 19-40. <https://www.lamjol.info/index.php/DIALOGOS/article/view/2202>
- Supo, J. (2013). *Cómo validar un instrumento. Aprende a crear y validar instrumentos como un experto*. http://cua.uam.mx/pdfs/coplavi/s_p/doc_ng/validacion-de-instrumentos-de-medicion.pdf
- The Joint Committee on Standards for Educational Evaluation. (2003). *The Student Evaluation Standards*. Corwin Press Inc. Thousand Oaks.
- Timmons, J., Podmostko, M., Bremer, C., Lavin, D. y Wills, J. (2005). Career planning begins with assessment: A guide for professionals serving youth with educational & career development challenges. ERIC. <https://files.eric.ed.gov/fulltext/ED485703.pdf>
- Ventura-León, J. y Caycho, T. (2017). El coeficiente omega: un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627. <https://www.redalyc.org/pdf/773/77349627039.pdf>
- Wilson, M. (2005). Constructing measures, an item response modeling approach. *Review of Constructing Measures*, 28(4). https://www.researchgate.net/publication/240362885_M_Wilson_Constructing_measures_An_item_response_modeling_approach_Erlbaum_Associates_Mahwah_NJ_2005
- Zegarra, L., Loza, C. y Pérez, V. (2011). Validación psicométrica del instrumento índice de función eréctil en pacientes con disfunción eréctil en Perú. *Rev Peru Med Exp Salud Publica*, 28(3), 477-83. <https://rpmesp.ins.gob.pe/index.php/rpmesp/article/view/526>

Fuentes de financiamiento

Autofinanciado.

Conflictos de interés

Ninguno.

Correspondencia

E-mail: herman.tarazona@udh.edu.pe