

VALIDEZ Y CONFIABILIDAD DE LOS INSTRUMENTOS
para trabajos de investigación

Dr. Walter Iván Abanto Vélez

28 de Junio de 2015

Elaborada por:

Dr. Walter Iván Abanto Vélez

Pswalterabanto@hotmail.com

Validez y confiabilidad de los instrumentos para trabajos de investigación

Para la recolección de datos son indispensables los instrumentos, los cuales deben precisar dos requisitos básicos: validez y confiabilidad. En tal sentido hablar de instrumentos de medición con procedimientos de validez y confiabilidad es hablar de objetividad y credibilidad de los resultados. Es decir no podemos pretender hacer un trabajo de **investigación** con carácter científico, si no contamos con instrumentos debidamente validados y fiables para la recogida de datos.

1.- VALIDEZ:



La validez está referida a la objetividad y claridad de los ítems o preguntas que se utiliza en un instrumento de medición. Es decir con que veracidad los ítems o preguntas miden lo que realmente se quiere medir. Por ejemplo, ante un instrumento para medir CREENCIAS (son presunciones básicas que han arraigado en la mente de las personas, por lo tanto es cognición o forma de pensar) podemos encontrar el siguiente ejemplo:

Ítems errado	Ítems correcto
Yo siento que la vida es una sola y hay que vivirla al máximo.	Yo opino que la vida es una sola y hay que vivirla al máximo.

Como se puede observar en el siguiente ejemplo si creencia (forma de pensar, opinar o modo de ver el mundo) está relacionada a la cognición, los ítems deben de estar direccionados a ese aspecto y no al sentir que contempla el área emocional.

Por otro lado la validez tiene tres tipos:

1.1.- Validez de Contenido: Grado en que un instrumento refleja un dominio específico del contenido de lo que se mide; sin omisiones y sin desequilibrios de contenido.

La validez de contenido se utiliza generalmente con instrumentos o test de rendimiento, y especialmente con los tests educativos y tests referidos al criterio. En este tipo de tests se trata de comprobar los conocimientos respecto a una materia o un curso.

Ejm: Una prueba de operaciones aritméticas no tendrá validez de contenido si incluye sólo problemas de adición y excluye problemas de sustracción, multiplicación y división.

1.1.1.- Validez de juicio de experto: Para Escobar Pérez, J. & Cuervo-Martínez, A. (2008) la validez de contenido generalmente se evalúa a través de un panel o un juicio de expertos. De igual modo el **juicio de expertos** se define como una opinión informada de personas con trayectoria en el tema, que son reconocidas por otros como expertos cualificados en éste, y que pueden dar información, evidencia, juicios y valoraciones. La identificación de las personas que formarán parte del juicio de expertos es una parte crítica en este proceso, frente a lo cual Escobar Pérez, J. & Cuervo-Martínez, A. (2008) cita a Skjong & Wentworht (2000) quienes proponen los siguientes criterios de selección:

- Experiencia en la realización de juicios y toma de decisiones basada en evidencia o experticia (grados, investigaciones, publicaciones, posición, experiencia y premios entre otras).
- Reputación en la comunidad.
- Disponibilidad y motivación para participar.
- Imparcialidad y cualidades inherentes como confianza en sí mismo y adaptabilidad.

También plantean que los expertos pueden estar relacionados por educación similar, entrenamiento, experiencia, entre otros; y en este caso la ganancia de tener muchos expertos disminuye. Otros

autores como McGartland, Berg, Tebb, Lee y Rauch (2003), proponen como criterio básico de selección únicamente el número de publicaciones o la experiencia.

De otra parte, el número de jueces que se debe emplear en un juicio depende del nivel de experticia y de la diversidad del conocimiento; sin embargo, la decisión sobre qué cantidad de expertos es la adecuada varía entre autores. Así, mientras Gable y Wolf (1993), Grant y Davis (1997), y Lynn (1986) (citados en McGartland et al. 2003) sugieren un rango de **2** hasta **20 expertos**, Hyrkäs et al. (2003) manifiestan que **10 expertos** brindarían una estimación confiable de la validez de contenido de un instrumento. Si un 80 % de los expertos han estado de acuerdo con la validez de un ítem éste puede ser incorporado al instrumento (Voutilainen & Liukkonen, 1995, citados en Hyrkäs et al. (2003).

Sin embargo teniendo en cuenta lo mencionado anteriormente y conociendo nuestra realidad sociocultural se recomienda que el número de expertos sea de **7 a más**.

TABLA DE CRITERIO PARA CONSIDERAR A UN EXPERTO O JUEZ

Grado Académico	Maestría 1 pto.		Doctorado 2 ptos.	
Experiencia en el área de estudio asociado a su variable	Clínica 1 pto.	Educativa 1 pto.	Organizacional 1 pto.	Social 1 pto.
Experiencia Laboral en el área	De 2 a 4 años 1 pto.		Más de 5 años 2 ptos.	
Experiencia en investigaciones Psicométricas	Trabajo(s) psicométricos realizados Título del estudio realizado(_____) 1 pto.			

Para poder considerar a un experto o juez en el proceso de validez se debe tener las siguientes consideraciones:

- El puntaje mínimo para aceptar a un experto o juez es de 4 puntos según la tabla de criterio.
- Los puntajes de cada criterio no son acumulables; por ejemplo: si tiene doctorado el cual equivale a 2 puntos ya no se puede considerar los puntajes de maestría.
- Debe obtener el puntaje mínimo en el grado académico, experiencia en el área de estudio asociado a la variable y experiencia laboral en el área. Respecto a la experiencia en investigaciones psicométricas puede ser opcional, los otros es obligatorio tener como mínimo 1 punto.
- La experiencia laboral en el área implica estar trabajando cómo psicólogo del área profesional de manera frecuente.

1.2.- Validez de Criterio: Un **criterio** es una variable distinta del test que se toma como referencia, que se sabe que es un indicador de aquello que el test pretende medir o que se sabe que debe presentar una relación determinada con lo que el test pretende medir.

En otras palabras la validez de criterio se establece al validar un instrumento de medición al compararlo con algún criterio externo que pretende medir lo mismo. Validez concurrente y la validez predictiva.

Ejemplo: En las campañas electorales, los sondeos se comparan con los resultados finales de las elecciones.

A la correlación entre las puntuaciones en el test (X) y en el criterio (Y) se le denomina coeficiente de validez. Dentro de la validez de criterio se pueden utilizar los estadísticos de Coeficiente de Contingencias, Spearman – Brow, Pearson, Alfa de Cronbach y la Técnica Aiken.

Según Casas Vilchis, Luis (2011) refiere que este tipo de validez se divide en concurrente y predictiva:

1.2.1.- Validez concurrente: Esta referido al grado en que un instrumento mide lo mismo que miden otros instrumentos ya existentes, considerados como medidas adecuadas del mismo rasgo. Así mismo se da cuando el instrumento se correlaciona con algún criterio existente que mida lo mismo que el instrumento (ejemplo: en un hospital se utiliza la producción del personal administrativo como criterio para evaluar a personal nuevo y como criterio para validación de una prueba para selección de personal administrativo). Es relevante precisar que la validez concurrente de un instrumento no indica nada acerca de la capacidad predictiva de éste, solo señala que el instrumento nuevo que se ha elaborado para medir un determinado atributo es tan eficiente o válida como el criterio previo.

1.2.2.- Validez predictiva: Hace referencia al grado en que la aplicación de un instrumento o test sirve para predecir la conducta futura de las personas que han sido evaluadas. Se deben señalar coeficientes de validez predictiva para cada tipo de predicción que se pretenda hacer con un mismo instrumento (ejemplo: exámenes de inteligencia para medir éxito académico de los sujetos en distintos grados escolares). La predicción del comportamiento futuro, con un instrumento que se administre en el presente, se limita al lapso incluido en la determinación de la validez predictiva del instrumento. Solo es válida su capacidad predictiva de acuerdo al lapso en que se ha probado. La validez predictiva es esencial y se requiere en las pruebas de aptitud e inteligencia. El psicólogo que hace predicciones sobre comportamiento futuro sin tener sustento para tales, incurre a una mala práctica sujeta a consideraciones éticas y legales.

1.3.- Validez de Constructo: El término constructo hace referencia a un concepto teórico psicológico inobservable (ej. la inteligencia, personalidad, las aptitudes, asertividad, etc.). Es decir no son directamente visibles.

Estos constructos a pesar de no ser distinguibles al ojo humano; a pesar de estar basado en teorías psicológicas, precisan de indicadores observables para su estudio. En tal sentido es necesario que estos indicadores estén precisados en los ítems de un instrumento o test, y debe comprobarse empíricamente que resultan adecuados para reflejar el constructo de referencia.

Así también la validez de constructo debe explicar el modelo teórico empírico que subyace a la variable de interés. Ejm: El Análisis de Factores y Análisis de Cofactores, el Análisis de Covarianza.

Para la estimación de la validez de constructo se utiliza una metodología variada. Algunos de los métodos más utilizados son:

1.3.1.- Análisis factorial del test: El análisis factorial permite ordenar los datos y facilitar la interpretación de las correlaciones. Se espera un factor explicativo del constructo con saturaciones altas del test y los tests que miden aspectos parecidos, y con saturaciones bajas de aquellos tests que miden aspectos diferentes. Con frecuencia se habla de la estructura factorial de un test como validez estructural o validez factorial.

Se utiliza para probar que rasgos técnicos (factores o componentes, constructos) se miden a través de los ítems del test o instrumento.

1.3.2.- Validez Convergente- Discriminante. Matrices Multirasgo – Multimétodo

Muy en relación con la validez factorial se encuentra también la validez convergente- discriminante, la cual se evalúa mediante las matrices multirasgo- multimétodo. El sentido de estas definiciones es el siguiente:

a) Validez convergente: Está referido a aplicar dos test que miden el mismo atributo y establecer la relación existente entre ambos resultados; vale aclarar que la correlación debe ser significativa o alta (relación positiva).

b) Validez discriminante: Indica que dos test miden atributos diferentes, en tal sentido sus correlaciones debe ser baja o nula.

A través de estos dos tipos de validez se podría ir definiendo un constructo psicológico.

1.3.3.- Análisis de las diferencias individuales que pone de manifiesto un test: Se refiere al análisis de la distribución de las puntuaciones de test y a comparaciones de estos aspectos en distintas muestras. Diferentes edades, sexos, niveles profesionales, etc. Estas comparaciones no son arbitrarias, sino que se derivan de hipótesis que se hacen en función de los conocimientos que se tiene del constructo.

1.3.4.- Análisis de los cambios en las diferencias individuales: Se refiere a la investigación diacrónica de los mismos sujetos con el mismo test. Este tipo de estudios permite conocer la estabilidad del rasgo a lo largo del tiempo y a través de situaciones.

1.3.5.- Análisis lógico de los elementos del test: Se refiere al análisis de ítems del test en relación con el constructo. Aquellos ítems que correlacionan positivamente entre sí, pertenecen al mismo constructo. Esto significa que el análisis de consistencia interna de un test no sólo aporta datos respecto a su fiabilidad, sino a su validez. Si se obtiene un coeficiente de consistencia interna bajo, significa que el test no mide un único constructo. Todos estos métodos enumerados se complementan entre sí. Se trata de ver el constructo que trata de medir el test desde diferentes ángulos. Por tanto, no existe una única medida de la validez de constructo.

2.- CONFIABILIDAD:

Menéndez, A. (s.f.) Refiere que la confiabilidad es la consistencia de los resultados. En el análisis de la confiabilidad se busca que los resultados de un cuestionario concuerden con los resultados del mismo cuestionario en otra ocasión. Si esto ocurre se puede decir que hay un alto grado de confiabilidad. También se habla de confiabilidad cuando dos o más evaluadores evalúan al mismo estudiante sobre el mismo material y se obtienen puntuaciones semejantes.

Por otro lado Arias Pittman, J. (2013) refiere que la confiabilidad del instrumento de medición constituye el criterio fundamental para evaluar su calidad y adecuación; no es un concepto unitario que pueda ser descrito o definido con facilidad, lo que constituye desventaja para el investigador novato. En esencia la confiabilidad de un instrumento es el grado de congruencia con la que mide el atributo que se supone evalúa. Si una balanza arroja un dato de 54.43k para el peso de una persona ni un minuto después da una lectura de 68.03k (descontando cualquier intromisión con el instrumento o el sujeto), debemos tener gran cautela al usarla porque la información que genera será poco fidedigna. Cuanto menor variación produzca un instrumento de mediciones repetidas de un atributo, mayor será su confiabilidad de este modo, la confiabilidad puede equiparse con la estabilidad congruencia o confiabilidad de un instrumento de medición.

Otra forma de definir la confiabilidad es en términos de precisión es posible señalar que un instrumento es confiable si sus mediciones reflejan exactamente las mediciones "verdaderas" del atributo en estudios. La definición anterior vincula la confiabilidad con puntos señalados en nuestro comentario del error de medición. Intentaremos esclarecer aún más tal medición, para indicar que el instrumento es confiable y preciso en la medida en que no existan errores de medición en los resultados obtenidos. En otras palabras, una medición confiable es aquella que lleva al máximo el componente verdadero y el mínimo componente de error. Cuando mayor el error, mayor será la falta de confiabilidad.

Por consiguiente podemos decir que la confiabilidad es la "exactitud o precisión de un instrumento de medición (Arias Pittman, J. 2013 p. 7)". Existen distintos tipos de confiabilidad:

- La estabilidad a través del tiempo (medible a través de un diseño test-retest);

- La representatividad, que se refiere a la ausencia de variaciones en la capacidad del instrumento para medir un mismo constructo en distintas subpoblaciones; y por último
- La equivalencia o formas paralelas, que se aplica a las variables latentes, medidas a través de múltiples indicadores, y que se puede poner a prueba mediante diversos métodos, incluyendo el llamado Alpha de Cronbach, split-half, y distintas formas de verificar la consistencia entre evaluadores.

Alfa de Cronbach:

El coeficiente alfa de Cronbach se basa en el cálculo de la confiabilidad de un compuesto donde cada ítem se considera una subcuestionario del cuestionario total y los ítems se consideran cuestionarios paralelos. Como esta propiedad de paralelismo es prácticamente imposible para los ítems, por lo general el coeficiente alfa de Cronbach subestima el coeficiente de correlación. $\alpha < CC$ • La fórmula para el coeficiente alfa es

$$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

k = número de ítems

$(\sigma_i)^2$ = varianza de cada ítem

$(\sigma_x)^2$ = varianza de la cuestionario total

3.- FACTORES QUE AFECTAN LA VALIDEZ Y LA CONFIABILIDAD DE UN INSTRUMENTO DE MEDICIÓN:

Hay diversos factores que pueden afectar la confiabilidad y la validez de los instrumentos de medición; algunos podrían ser:

- 3.1.-** Improvisación; pretender elaborar un instrumento de recolección de información de un día para otro o utilizar una prueba sin ningún criterio técnico, metodológico o científico que se ajuste a las necesidades de lo que se pretende investigar.
- 3.2.-** Utilización de instrumentos desarrollados en el extranjero que no han sido sometidos al proceso psicométrico de validez y confiabilidad en el propio contexto. Las culturas según su nación o región son diferentes; y esto debemos tomarlo en cuenta al elegir un instrumento de medición.
- 3.3.-** Inadecuación o incompatibilidad para las personas a las que se aplica; es decir el instrumento está diseñado para ser aplicado a mayores de 25 años y se aplica a personas de menor edad consignada en la prueba.
- 3.4.-** Condiciones en la que se aplica; es decir su aplicación se realiza con muchos estímulos distractores o en lugares inapropiados e incluso cuando la persona evaluada se encuentra apurado/a en querer retirarse, generando que rinda su evaluación rápidamente marcando o respondiendo a las preguntas de forma subjetiva.
- 3.5.-** Pobreza en el N° de sujetos que se utiliza para determinar la validez y confiabilidad de un instrumento; es decir pretender validar y determinar la fiabilidad de un instrumento con una muestra de 20 o 30 personas. En tal sentido se recomienda establecer estadísticamente una muestra representativa o utilizar un criterio con rigor científico.
- 3.6.-** Número de ítems de la prueba supera el número de muestra para establecer la validez y fiabilidad; es decir el instrumento que se pretende utilizar tiene 80 ítems y la muestra para validar es de 20 sujetos.

REFERENCIAS:

Arias Pittman, J. (2013). Confiabilidad y Validez para evaluar los instrumentos de medición.
Recuperado de:

<https://docs.google.com/document/d/10ECoSNkohHF50qE1s1axrujaDawaNPYG2iv677tCKJs/edit>

Casas Vilchis, L. (2011). Elaboración de Instrumentos (Validez de Criterio)

<https://luiscasavilchis.wordpress.com/2011/09/15/elaboracion-de-instrumentos-validez-de-criterio/>

Chiner Esther (S./F.). La Validez. Tema 6. Recuperado de:

<http://rua.ua.es/dspace/bitstream/10045/19380/25/Tema%206-Validez.pdf>

Escobar Pérez, J. & Cuervo-Martínez, A. (2008). VALIDEZ DE CONTENIDO Y JUICIO DE EXPERTOS: UNA APROXIMACIÓN A SU UTILIZACIÓN. Avances en Medición, 6, 27–36. Universidad El Bosque, Colombia. Recuperado de:

http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf

Menéndez, A. (S. / F.) Confiabilidad. Taller CES revisado en la página. Recuperado de:

<http://www.gobierno.pr/NR/rdonlyres/CC1286A8-310F-48CF-AB2C-D30417D9AF78/0/15confiabilidad.pdf>