

Metodología para la validación de instrumentos de medición.

Methodology for the validation of measuring instruments.

DI. Merly Rosa Nápoles Fernández

merlyrnf@gmail.com

ORCID: 0000-0003-1556-6958

Instituto Superior de Diseño

Universidad de La Habana

Cuba

Autor para correspondencia

Dra. C. Noelia Barrueta Gómez

noeliab@isdi.co.cu

ORCID: 0000-0002-0313-4376

Instituto Superior de Diseño

Universidad de La Habana

Cuba

RESUMEN

En el contexto de la investigación científica, la medición es un asunto de vital relevancia para cuantificar los aspectos que competen a la misma. En este proceso, el instrumento de medición tiene un papel central, si este no representa a las variables de estudio adecuadas, la investigación no es digna de tomarse en cuenta. Estudios sobre este tema abarcan ampliamente sus diferentes cuestiones de manera dispersa y no describen de forma minuciosa los detalles del proceso. De ahí que este trabajo persiga brindar a los profesionales y estudiantes del Instituto Superior de Diseño (ISDi), información detallada y una metodología a seguir que facilite este procedimiento. De esta manera, se recogen los diferentes aspectos que debe presentar un instrumento de medición adecuado, específicamente un cuestionario y detalladamente, aquello que involucra su validación.

ABSTRACT

In the context of scientific research, measurement is a matter of vital relevance in quantifying the aspects within it. In this process, the measurement instrument plays a central role, if it does not represent the appropriate study variables, research is not relevant. Studies on the subject broadly cover their different issues at a high level and do not describe thoroughly the details of the process. Hence, this work pursues to provide ISDi professionals and students with detailed information and a thorough methodology to follow that facilitates this procedure. This methodology includes different aspects that an appropriate measuring instrument, specifically a questionnaire, must present and in detail what involves its validation.

Palabras claves:

validación
cuestionario
validez
confiabilidad
objetividad

Keywords:

validation
questionnaire
validity
reliability
objectivity

Fecha Recibido:

07 / 03 / 2022

Fecha Aceptación:

10 / 05 / 2022

Fecha Publicación:

01 / 01 / 2023

INTRODUCCIÓN

El instrumento de medición es un recurso que utiliza el investigador para recolectar y registrar información o datos sobre las variables que tiene en mente. Los académicos reconocen que la aplicación de un enfoque inadecuado de la medición en un estudio puede generar datos inapropiados, de ahí la importancia de desarrollar instrumentos de medición adecuados.

Para garantizar el desarrollo de estos en las investigaciones que así lo requieran desarrolladas en el Isdi se ofrece de manera concisa un acercamiento a los aspectos a considerar, así como a procedimientos que pueden ser utilizados, en dependencia de las características de la investigación.

Debe tenerse en cuenta que la construcción y validación de un instrumento de medición es un proceso complejo, del que en el presente trabajo se destacan algunos de los elementos más relevantes a fin de orientar el desarrollo del mismo.

En la investigación se dispone de diversos tipos de instrumentos para medir las variables de interés y en algunos casos llegan a combinarse varias técnicas de recolección de los datos. Estos pueden: cuestionarios, escalas de actitudes, registros del contenido (análisis de contenido) y observación, pruebas estandarizadas (medidas del desempeño individual), recolección de información factual e indicadores (análisis de datos secundarios de registros públicos y documentación) y meta-análisis.

El cuestionario constituye uno de los instrumentos más utilizados para la recolección de datos. Este consiste en un conjunto de preguntas respecto a una o más variables a medir, siendo congruente con el planteamiento del problema o hipótesis (Hernández Sampieri et al., 2010). Existen numerosos estilos y formatos de cuestionarios, de acuerdo a la finalidad específica de cada uno de ellos (Nieto, 2011). Estos representan un medio útil y eficaz para obtener información de los consultados en un tiempo relativamente breve, por lo que resultan vitales para recabar datos de grupos numerosos.

No obstante, su construcción, aplicación y tabulación poseen un alto grado científico y objetivo por lo que elaborar un cuestionario válido no es una cuestión sencilla; implica controlar una amplia gama de variables, las cuales se deberán comprobar durante la validación para verificar que el instrumento cumple con los aspectos necesarios.

Se llevan a cabo los métodos teóricos histórico-lógico y análisis y síntesis.

RESULTADOS Y DISCUSIÓN:

Partiendo del estudio de la bibliografía consultada y la sistematización de la misma, se recogen los diferentes elementos con los que debe contar un instrumento de medición, específicamente un cuestionario, de modo que estos sean verificados en el proceso de validación del mismo. De esta forma se aclaran conceptos básicos asociados a la temática para un mejor entendimiento de la misma, descomponiéndola en los diferentes factores que la componen: validez, validez de contenido, validez de constructo, validez de criterio, confiabilidad y objetividad. Se presenta una metodología a seguir para la validación de un instrumento, detallando diferentes métodos utilizados para cada aspecto, así como los procedimientos necesarios y la final interpretación de los resultados que se obtengan.

DISCUSIÓN

Atendiendo a las recomendaciones de Hernández Sampieri et al. (2010), cuando se construye un cuestionario deben considerarse diferentes cuestiones básicas, las cuales deberán ser verificadas en su validación. Por este motivo a continuación se recogen aspectos esenciales a analizar durante este proceso.

En primer lugar, un cuestionario deberá estar compuesto por portada, introducción, instrucciones, datos del encuestado y agradecimiento final, además de las preguntas y categorías de respuestas. En algunos casos se incluye también la constancia del participante y su autorización para utilizar los datos proporcionados. En el caso de los formatos de cuestionarios, estos agregan una hoja de respuestas, para facilitar el procesamiento de los datos.

En la preparación del cuestionario debe quedar claro cómo se interpretarán las respuestas obtenidas, cuantitativa y cualitativamente. Para tal fin, es necesario que la variable de investigación, sus dimensiones, indicadores e ítems queden debidamente determinados y codificados.

En cuanto a las preguntas, estas deberán basarse en los objetivos del cuestionario, habiendo quedado estos claramente definidos desde el inicio del estudio. Estas deben estar redactadas de forma coherente, organizada, secuenciada y estructurada; ser objetivas, claras, precisas y comprensibles para los sujetos encuestados, evitando términos confusos, ambiguos, de doble sentido o con múltiples significados. Las mismas no pueden hacer supuestos acerca de los encuestados, ni ser percibidas como amenazantes. Tampoco deberán inducir una respuesta o inclinar hacia la misma, así como no es aconsejable que abarquen varios temas en una misma interrogante. Sus categorías de respuesta deben ser exhaustivas, con un balance de alternativas positivas y negativas y de ser posible, una alternativa neutra.

El cuestionario debe ser y parecer corto, fácil y atractivo. No existe una regla respecto a la longitud que debe tener un cuestionario, de ser muy corto se pierde información y de resultar largo puede llegar a ser tedioso. Por tanto, el tamaño depende del número de variables y dimensiones a medir, el interés de los participantes y la manera como se administre. Cuestionarios que duran más de 35 minutos suelen resultar fatigosos, a menos que los sujetos estén muy motivados para contestar.

De utilizar un cuestionario diseñado anteriormente en su totalidad, se requiere el permiso del autor y de conseguirse, otorgarle crédito en la referencia correspondiente.

En sentido general, toda medición o instrumento de recolección de datos debe reunir tres requisitos esenciales: confiabilidad, validez y objetividad (Hernández Sampieri et al., 2010). A la comprobación de estos elementos en un instrumento de medición es a lo que corresponde la validación del mismo. Los cuestionarios

adquiridos no quedan exentos de validación, estos requisitos deberán comprobarse bajo condiciones locales. Es por ello que resulta vital analizar cada uno de estos aspectos y recoger los métodos y procedimientos que posibilitan su comprobación.

Validez

Según Hernández Sampieri et al. (2010) la validez de un test indica el grado de exactitud con el que mide el constructo teórico que pretende medir y si se puede utilizar con el fin previsto.

Sobre este concepto pueden tenerse diferentes tipos de evidencia, dígase, evidencia relacionada con el contenido, evidencia relacionada con el criterio y evidencia relacionada con el constructo (Babbie, 2009).

La validez de un instrumento de medición se evalúa sobre la base de todos los tipos de evidencia. Cuanta mayor evidencia de validez de contenido, de validez de criterio y de validez de constructo tenga un instrumento de medición, este se acercará más a representar la(s) variable(s) que pretende medir (Hernández Sampieri et al. (2010). Sin embargo, otros autores plantean que las tres se refieren a aspectos diferentes y la utilización de uno u otro concepto de validez depende del tipo de test.

Validez de contenido

La validez de contenido de un instrumento se refiere al grado en que el test presenta una muestra adecuada de los contenidos a los que se refiere, sin omisiones y sin desequilibrios de contenido (Nieto, 2011). Un instrumento de medición requiere tener representados prácticamente a todos o la mayoría de los componentes del dominio de contenido de las variables a medir. “La pregunta que se responde con la validez de contenido en un cuestionario, sería: ¿las preguntas qué tan bien representan a todas las preguntas que pudieran hacerse?” (Hernández Sampieri et al., 2010).

Métodos para determinar la validez de contenido.

Según Sireci (2003), se pueden establecer dos planteamientos para estimar la validez de contenido: métodos basados en el juicio de expertos y la utilización de métodos estadísticos derivados de la aplicación del instrumento de medida.

Los métodos basados en el juicio de expertos se caracterizan por contar con un número de expertos que bien proponen los ítems o dimensiones que deben conformar el constructo de interés o evalúan los diferentes ítems.

La apropiada selección de los expertos supone una cuestión fundamental a la hora de establecer este tipo de validez. Por ello, resulta fundamental analizar las características y experiencia de los expertos en relación al constructo tratado (Lawshe, 1975), para lo que se recomienda la metodología propuesta por el Comité Estatal para la Ciencia y la Técnica de Rusia, elaborada en 1971.

En primera instancia se confecciona un listado inicial de personas que posiblemente cumplan los requisitos para ser expertos en la materia a trabajar, previamente consultada su disposición para participar. Luego de confeccionar este listado, a los seleccionados se les somete a una autovaloración de los niveles de información y argumentación que poseen sobre el tema en cuestión para lo cual se aplica el formato de cuestionario propuesto por dicha metodología, el cual puede ser consultado en la referencia bibliográfica Nápoles, 2020. Finalmente son elegidos teniendo en cuenta los resultados arrojados por el cuestionario y otros criterios de selección como el nivel de conocimientos y capacidad predictiva, el grado de afectación por las consecuencias del proyecto, su capacidad facilitadora y el grado de motivación.

Luego de seleccionado el grupo de expertos a encuestar, se procede a preparar el cuestionario o guía de valoración para compilar sus apreciaciones sobre los diferentes elementos del instrumento de medición, de modo tal que el procesamiento de sus criterios y opiniones pueda realizarse a partir de los diferentes métodos y coeficientes utilizados para tal fin. Este formato de evaluación, al ser un instrumento de medición

en sí, debe cumplir con las características plasmadas anteriormente.

La valoración de los expertos suele realizarse en base a una escala tipo Likert. Estas escalas pueden presentar ligeras modificaciones, bien en cuanto al número de alternativas empleadas, las propuestas varían entre las cinco alternativas (Mussio & Smith, 1973) y las tres planteadas por Hambleton (1980), o bien en cuanto a la tarea en sí, solicitando valorar aspectos como la utilidad, relevancia, importancia, entre otras, de cada elemento. En cuanto a criterios de valoración no hay un estándar al respecto, por su parte Nieto (2011) recomienda los siguientes: pertinencia (grado de correspondencia entre el enunciado del ítem y lo que se pretende medir), claridad conceptual (hasta qué punto el enunciado del ítem no genera confusión o contradicciones), redacción (si la sintaxis, ortografía y la terminología utilizadas son apropiadas) y escala y codificación (si la escala empleada en cada ítem es apropiada y la misma ha sido debidamente codificada). Por otro lado, propone diferentes formatos de evaluación, en dependencia de los objetivos y características de la investigación.

Dado que la falta de independencia de los expertos puede constituir un inconveniente a la hora de evaluar un instrumento, es recomendable entregar la guía de valoración y el instrumento a evaluar personalmente para una autoadministración y valoración individual. Esta técnica se caracteriza por permitir el análisis de un problema complejo dando independencia y tranquilidad a los participantes, es decir, a los expertos.

Al margen del análisis cualitativo de los expertos, resulta imprescindible que estos aporten una valoración cuantitativa a los ítems. Por esta razón, es fundamental aplicar alguno de los métodos empíricos existentes para cuantificar este grado de acuerdo. Los métodos propuestos para determinar la validez de contenido que han tenido una mayor difusión y aplicación a nivel práctico son los siguientes:

1- Método de Lawshe. Índice de Validez de contenido (IVC). (Lawshe, 1975): Este método consiste en

la evaluación individual de los ítems de un test por parte de un grupo de expertos en la materia. A continuación, mediante la Razón de Validez de Contenido, se determina qué ítems del instrumento son adecuados y deben mantenerse en la versión final del mismo. En este punto, se debe asignar a cada ítem una puntuación en base a tres posibilidades: que el elemento sea esencial para evaluar el constructo, que resulte útil, pero prescindible o que se considere innecesario. Sobre esta valoración se aplica la siguiente expresión:

$$RVC = \frac{n - \frac{N}{2}}{\frac{N}{2}}$$

Donde n es el número de expertos que otorgan la calificación de esencial al ítem y N , el número total de expertos que evalúan el contenido. Finalmente, se calcula el Índice de Validez de Contenido (IVC) para el instrumento en su conjunto, el cual no es más que un promedio de la validez de contenido de todos los ítems seleccionados en el paso previo.

En cuanto a la interpretación de este índice el propio Lawshe elaboró una tabla que relaciona los valores obtenidos en este índice y el número de expertos empleado. De este modo, el valor crítico de la RVC se incrementa de manera monótona cuando se emplean entre 40 y 9 expertos (siendo los valores mínimos adecuados de 0.29 y 0.78, respectivamente) y alcanzando el grado máximo de acuerdo (0.99) cuando se recurre a 7 expertos o menos.

2- Coeficiente V de Aiken (Aiken, 1985): Este coeficiente puede ser calculado a partir de las valoraciones de un conjunto de expertos con relación a un ítem o conjunto de estos (Escrura Mayaute, 1988). El mismo se computa como la razón de un dato obtenido sobre la suma máxima de la diferencia de los valores posibles mediante la siguiente fórmula:

$$V = \frac{S}{(n(c - 1))}$$

Siendo: S = la sumatoria de S_i ; S_i = valor asignado por el experto i ; n = número de expertos y c = número de valores de la escala de valoración.

El coeficiente resultante puede tener valores entre 0 y 1. Cuanto más el valor computado se acerque a 1, entonces el ítem tendrá una mayor validez de contenido. Un adecuado valor del índice de acuerdo depende del tamaño de la muestra de expertos que se estudia, por lo que se recomienda evaluar el resultado estadísticamente haciendo uso de la tabla de probabilidades asociadas de cola derecha tabuladas por el autor y aceptar como válidos solo los ítems que sean estadísticamente significativos a $p < 0.05$. Dicho procedimiento permite calcular con un nivel de probabilidad la adecuación del contenido de los grupos de ítems atendiendo a cada grupo de expertos participantes. Es precisamente esta posibilidad de evaluar su significación estadística lo que hace a este coeficiente uno de los más apropiados para estudiar este tipo de validez.

3- Coeficiente de Validez de Contenido (CVC) (Nieto, 2002): La propiedad fundamental del Coeficiente de Validez de Contenido (CVC), es que además de medir la validez de contenido, este mide simultáneamente el grado de confiabilidad entre los jueces. Este permite valorar el grado de acuerdo de los expertos (el autor recomienda la participación de entre tres y cinco expertos) respecto a cada uno de los diferentes ítems y al instrumento en general. Para ello, tras la aplicación de una escala tipo Likert de cinco alternativas, se calcula la media obtenida en cada uno de los ítems y, en base a esta, se calcula el CVC para cada elemento siguiendo la fórmula siguiente:

$$CVC_i = \frac{M_x}{V_{max}}$$

Donde M_x representa la media del elemento en la puntuación dada por los expertos y V_{max} la puntuación máxima que el ítem podría alcanzar. Por otro lado, debe calcularse el error asignado a cada ítem (Pe_i), de este modo se reduce el posible sesgo introducido por alguno de los jueces, obtenido mediante la ecuación siguiente:

$$Pe_i = \left(\frac{1}{j}\right)^j$$

Siendo j el número de expertos participantes. Finalmente, el CVC se calcularía aplicando la fórmula:

$$CVC_t = \frac{\sum CVC_N}{N} = \sum \left[\left[\frac{\sum x_i}{J} \right] - P_{EI} \right] \left(\frac{1}{N} \right)$$

Esta quedaría simplificada de la siguiente manera:

$$CVC = CVC_i - P_{ei}$$

Respecto a su interpretación, la escala evaluativa del CVC plantea que con valores de 0 a 0.40 esta es inaceptable, de 0.41 a 0.60 muy baja, de 0.61 a 0.70 baja, de 0.71 a .79 moderadamente baja, de 0.80 a 0.90 buena y de 0.91 a 1.00 excelente. Nieto (2002) recomienda mantener únicamente aquellos ítems con un CVC superior a 0.80, aunque algunos criterios menos estrictos establecen valores superiores a 0.70 (Balbinotti, 2004).

Validez de constructo

Ya que un constructo es una conceptualización teórica sobre un aspecto medible del comportamiento, la validez de constructo trata de evaluar hasta qué punto una prueba mide los constructos sobre los que se sustenta. Se trata de comprobar si el instrumento mide el rasgo o concepto teórico o si se cumplen las hipótesis sobre la estructura del constructo (Martínez Arias, 1995). Bajo el criterio de Grinnell, Williams y Unrau (2009), esta es probablemente la más importante, sobre todo desde una perspectiva científica.

El proceso de validación de un constructo está vinculado con la teoría. Cuanto más elaborada y comprobada se encuentre la teoría que apoya la hipótesis, la validación del constructo arrojará mayor luz sobre la validez general de un instrumento de medición (Hernández Sampieri et al.2010).

Métodos para determinar la validez de constructo

Entre los enfoques o procedimientos comúnmente utilizados para establecer la validación de constructo se pueden citar: el método lógico, el método correlacional y el método experimental. Dentro de los procedimientos o técnicas estadísticas utilizados para la contrastación de la validez de constructo mediante el método correlacional destaca en mayor medida el análisis factorial, siendo la técnica por excelencia utilizada para tal fin.

Existen dos tipos de técnicas de análisis factorial conocidas: el análisis factorial confirmatorio (AFC) y el análisis factorial exploratorio (AFE).

En el AFC, “el investigador no sólo tiene una hipótesis previa sobre la existencia de factores comunes, también tiene una hipótesis previa sobre el número de factores” (Álvarez, 1994). Dadas estas condiciones, el análisis factorial confirmatorio puede confirmar o no las hipótesis, pero este no debe utilizarse sin una buena justificación. Autores plantean que un criterio razonable para utilizar el análisis factorial confirmatorio es que el instrumento no sea de nueva creación y que se conozca ya la estructura factorial obtenida en otras muestras. Dicho análisis requiere pruebas complementarias de bondad de ajuste, para confirmar si la estructura obtenida coincide con la estructura propuesta como hipótesis.

Por otro lado el objeto del AFE es “establecer una estructura subyacente entre las variables del análisis, a partir de estructuras de correlación entre ellas; o, en otras palabras: busca definir grupos de variables (más conocidos como factores) que estén altamente correlacionados entre sí” (Méndez y Rondón, 2012). El análisis factorial exploratorio es considerado más un procedimiento para generar teorías que para confirmarlas, sin embargo, este es el utilizado con mayor frecuencia, aunque la intención sea confirmatoria, cuando las especificaciones propias del análisis factorial confirmatorio no están seriamente justificadas, práctica que algunos autores plantean no es siempre la más recomendable. Algunos investigadores hacen en primer lugar un análisis exploratorio y confirman después la estructura factorial con un análisis factorial confirmatorio en una nueva muestra, lo cual se considera una buena práctica según Henson y Roberts, 2006 citado por Morales, 2011.

A partir de los análisis de Álvarez (1995) y Hair y colaboradores (citados por Méndez et al., 2012) sobre los pasos a tener en cuenta al realizar el AFE se observa que en primer lugar es importante definir los objetivos que persigue su ejecución. Posteriormente se realiza el examen de la matriz de correlaciones que consiste en analizar la existencia de altas correlaciones entre

las variables; lo cual es indicativo de información redundante. En este paso es conveniente realizar una serie de pruebas que indiquen si es pertinente desde el punto de vista estadístico el AFE con los datos de la muestra disponible. Entre las principales pruebas tenemos: determinante de la matriz de correlaciones; esfericidad de Barlett e índice Kaiser, Meyer y Olkin (KMO). En este sentido Méndez et al. (2012), recomiendan “aplicar al menos dos de las medidas anteriormente mencionadas, y si alguna de ellas evidencia algún grado de correlación, se considera que tiene sentido realizar el análisis”.

Al determinar que el AFE es pertinente, se procede como tercer paso a la extracción de los factores, cuya finalidad es precisar las variables latentes que representan a las variables originales del estudio. Aunque existen distintos métodos para extraer los factores, los más usados son el método de componentes principales y el método de factor común.

El método de componentes principales busca los componentes que expliquen la mayor cantidad de varianza explicada (diferencias en las respuestas dadas en el instrumento) por factor en la matriz de correlaciones.

El método de factor común se fundamenta en la varianza compartida entre las variables; las principales desventajas de este método es que no tiene una única solución para los puntajes del factor, y cuando su varianza compartida no es estimable, es necesario eliminar variables del análisis (Méndez et al., 2012). Sin embargo, si el objeto es elegir las variables con pesos factoriales mayores en cada factor, no suele haber diferencias entre los dos tipos de métodos. Entre las diversas modalidades del análisis de Factores Comunes, las más recomendadas son las denominadas Maximum Likelihood si las distribuciones son aproximadamente normales y Principal Axis Factors si las distribuciones se apartan notablemente de la normal (Costello y Osborne, 2005).

A menudo es difícil interpretar los factores iniciales identificados en el paso anterior, por consiguiente, a la extracción inicial se le realiza la rotación con la finali-

dad de lograr una solución que facilite la interpretación. Existen dos sistemas de rotación: el método de rotación ortogonal (varimax, quartimax y equamax) y el método de rotación oblicuo (oblimin y promax).

Establecida la rotación se hace útil evaluar las ponderaciones que reciben las variables en los factores. En este último paso se hace “una evaluación tanto en términos estadísticos como prácticos, con el fin de determinar cuáles variables son las más importantes, y de la misma forma, identificar aquellas que definitivamente no aportan y pueden ser eliminadas del análisis” (Méndez et al., 2012). Estos mismos autores definen que las puntuaciones se interpretan en valores absolutos de la siguiente forma: valores menores a 0,3 se consideran no significativos; aporte mínimo entre 0,3 y menor a 0,5; aporte significativo entre 0,5 y 0,7; y valores mayores a 0,7 se consideran relevantes.

Los factores se consideran bien definidos cuando al menos tres variables tienen en él sus mayores pesos (Costello y Osborne, 2005). Si concebimos un factor como un constructo subyacente que explica las correlaciones entre los ítems, no tiene sentido hablar de un factor formado por un solo ítem: esto querría decir que ese ítem no se relaciona de manera especial con los demás factores (o grupos de variables). Al menos hacen falta dos variables para poder hablar de algo común. Un factor sólido vendría definido por unos 5 ítems con pesos de .50 o más en el factor.

Para realizar estos análisis, no existe un criterio o norma definitiva sobre el número de sujetos necesario como muestra a la que aplicar el instrumento. Algunos autores sugieren que el número de sujetos sea el doble que el número de variables, otros recomiendan utilizar una muestra 10 veces mayor que el número de variables o ítems ($N = 10k$; Nunnally, 1978; Thorndike, 1982). Una orientación más segura es que el número de sujetos no baje de 200 y que al menos haya 5 sujetos por ítem. En una revisión de 60 análisis factoriales de Henson y Roberts (2006), la media es de 11 sujetos por variable.

Validez de criterio

El principio de la validez de criterio consiste en que, si diferentes instrumentos o criterios miden el mismo

concepto o variable, deben arrojar resultados similares. Bostwick y Kyte (2005) (citado por Hernández Sampieri et al., 2010) expresan que, si hay validez de criterio, las puntuaciones obtenidas por ciertos individuos en un instrumento deben estar correlacionadas y predecir las puntuaciones de estas mismas personas logradas en otro criterio. Cuanto más se relacionen los resultados del instrumento de medición con el criterio, la validez de criterio será mayor. La elección del criterio es el aspecto crítico en este procedimiento de determinación de la validez.

Dentro del concepto de validez de criterio cabe distinguir a su vez entre validez externa y validez interna, según el criterio con que se evalúe y entre validez concurrente y validez predictiva, según la dimensión temporal que se establezca entre la prueba y el criterio.

Métodos para determinar la validez de criterio

Los procedimientos estadísticos utilizados en la validación referida a un criterio varían según el número de predictores utilizados (uno o más tests) y el número de criterios empleados (criterio único y criterio compuesto o múltiple). Martínez Arias (1995) distingue que cuando se emplea un único test y un solo criterio, se emplearían los procedimientos de correlación y regresión lineal simple. De tratarse de varios predictores (tests) y un solo criterio, se emplearía la correlación y regresión lineal múltiple o el análisis discriminante. Por último, cuando se utilicen varios predictores y varios criterios se emplea la regresión lineal multivariante y la correlación canónica. Estos procedimientos es posible realizarlos mediante los programas de análisis estadísticos como SPSS. La disponibilidad y facilidad de uso del software que permite la construcción de modelos de regresión ha hecho olvidar que se trata de técnicas complejas, que requieren un cierto conocimiento de la metodología estadística subyacente (Molinero Casares, 2002).

Confiabilidad

La confiabilidad de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo individuo u objeto produce resultados iguales, consistentes y coherentes (Hernández Sampieri et al., 2010).

En el sentido más amplio, la confiabilidad de una prueba indica el grado en que las diferencias individuales en las calificaciones de una prueba son atribuibles al error aleatorio de medición y en la medida en que son atribuibles a diferencias reales en la característica o variable que se está midiendo.

Métodos para determinar la confiabilidad

Los procedimientos más utilizados para determinar la confiabilidad mediante un coeficiente son la medida de la estabilidad, el método de formas alternativas o paralelas, el método de mitades partidas y la medida de la consistencia interna.

La ventaja de las medidas de consistencia interna es que requieren una sola administración del instrumento de medición. Estas son coeficientes que estiman la confiabilidad, dígase el alfa de Cronbach (desarrollado por J. L. Cronbach) y los coeficientes KR-20 y KR-21 de Kuder y Richardson (1937). El coeficiente alfa es para intervalos y los coeficientes Kuder Richardson para ítems dicotómicos. La mayoría de los programas estadísticos como SPSS y Minitab los determinan y solamente deben interpretarse.

Todos estos procedimientos producen coeficientes de fiabilidad que pueden oscilar entre cero y uno, donde un coeficiente de cero significa nula confiabilidad y uno representa un máximo de confiabilidad (fiabilidad total, perfecta).

El índice se puede generar por preguntas individuales y las que no cumplan con el índice requerido se eliminarán del estudio. Una cuestión importante es que los coeficientes son sensibles al número de ítems o reactivos, entre más ítems, el valor del coeficiente tenderá a ser más elevado.

Respecto a la interpretación del coeficiente cabe señalar que no hay una regla que indique que a partir de cierto valor no hay fiabilidad del instrumento. Más bien, el investigador calcula su valor, lo reporta y lo somete a escrutinio de los usuarios del estudio u otros investigadores. Sin embargo, en sentido general se espera que dicho índice esté por encima de 0,70, al ser este el valor de referencia y lo que permite concluir que el instrumento utilizado es válido. Concretamente

son valores aceptables de alfa para propósitos de investigación $\geq 0,7$ (aceptable), 0.80 (bueno) y para propósito de toma de decisiones $\geq 0,9$ (excelente).

Objetividad

La objetividad del instrumento se refiere al grado en que el instrumento es permeable a la influencia de los sesgos y tendencias de los investigadores que lo administran, califican e interpretan (Hernández Sampieri et al., 2010).

Esta se refuerza mediante la estandarización en la aplicación del instrumento (mismas instrucciones y condiciones para todos los participantes) y en la evaluación de los resultados; así como al emplear personal capacitado y experimentado en el instrumento. Por ejemplo, si se utilizan observadores, su proceder en todos los casos debe ser lo más similar que sea posible y su entrenamiento tendrá que ser profundo y adecuado.

Los estudios cuantitativos buscan que la influencia de las características y las tendencias del investigador se reduzca al mínimo posible, lo que representa es un ideal, pues la investigación siempre es realizada por seres humanos.

Procesamiento de resultados

Los resultados obtenidos mediante el procesamiento de los datos proporcionados por los expertos, aportan información relevante sobre la apreciación de los mismos para la validación del instrumento. Estos pueden ser agrupados según las partes que componen la guía de valoración entregada a los expertos, para la interpretación de las mismas a partir de las medidas de tendencia central, media, desviación estándar, mediana y moda. De igual forma, las sugerencias y recomendaciones plasmadas por los expertos deben ser procesadas y analizadas, lo cual unido a las calificaciones otorgadas por los expertos y los resultados de los coeficientes para cada uno de los diferentes análisis, permiten identificar aspectos de principal dificultad. Cada uno de estos elementos, así como los ítems propuestos a eliminación, deberán ser sujetos a revisión en el perfeccionamiento del instrumento para la consolidación de su versión final.

CONCLUSIONES:

1. Toda medición o instrumento de recolección de datos debe reunir tres requisitos esenciales: confiabilidad, validez y objetividad. En el caso de la validez esta puede tener diferentes tipos de evidencia, dígase, evidencia relacionada con el contenido, con el criterio y con el constructo.
2. Para estimar la validez de contenido pueden utilizarse métodos basados en el juicio de expertos o métodos estadísticos derivados de la aplicación del instrumento de medida.
3. Entre los enfoques o procedimientos comúnmente utilizados para establecer la validación de constructo se pueden citar: el método lógico, el método correlacional y el método experimental. El Análisis Factorial constituye uno de los aspectos más utilizados del enfoque correlacional.
4. Los procedimientos estadísticos más utilizados en la validación referida a un criterio son la correlación y regresión lineal, múltiple o canónica y el análisis discriminante.
5. Los procedimientos más utilizados para determinar la confiabilidad mediante un coeficiente son la medida de la estabilidad, el método de formas alternativas o paralelas, el método de mitades partidas y la medida de la consistencia interna.
6. Con excepción de los métodos de juicio de expertos para determinar la validez de contenido, los métodos para llevar a cabo la validación de un instrumento de medición requieren una aplicación previa del mismo a una muestra representativa, dígase la validez de constructo, validez de criterio y confiabilidad.
7. La objetividad se refuerza mediante la estandarización en la aplicación del instrumento y en la evaluación de los resultados; así como al emplear personal capacitado y experimentado en el instrumento.

REFERENCIAS BIBLIOGRÁFICAS:

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, pp. 131-142.

- Babbie, E. (2009). *The Practice of Social Research*. Cengage Learning.
- Costello, A. B., & Osborne, J. W. (2005). *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis*.
- Escorra Mayaute, L. M. (1988). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología*.
- Grinnell, R. M., Williams, M., & Unrau, I. (2009). *Research Methods for Social Workers*.
- Hambleton, R. K. (1980). Test score validity and standard setting methods.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. d. (2010). *Metodología de la Investigación*. Mexico D.F: McGraw-Hill.
- Lawshe. (1975). *A quantitative approach to content validity*.
- Martínez Arias, R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*.
- Méndez, C., & Rondón, M. (2012). Introducción al Análisis Factorial Exploratorio. *Revista Colombiana de Psiquiatría*.
- Molinero Casares, L. M. (2002). *Correlación y Regresión*.
- Morales Vallejo, P. (2011). *El Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios*. Madrid.
-