

METODOLOGÍA PARA LA VALIDACIÓN DE UNA ESCALA O INSTRUMENTO DE MEDIDA. Ramón Eugenio Paniagua Suárez. Profesor de la Universidad de Antioquia. Facultad Nacional de Salud Pública. Medellín, Colombia, 2015.

## Introducción

El método científico (1) pretende la adquisición de conocimientos, intenta resolver preguntas. Una vez formuladas las preguntas, se elaboran posibles respuestas fundamentadas en los conocimientos aceptados como consistentes por la comunidad científica; estas posibles respuestas se llaman hipótesis. Después de formuladas las hipótesis se puede diseñar un experimento o un estudio no experimental para contrastar las hipótesis. Poper decía que la ciencia busca explicaciones cada vez mejores. La secuencia del método científico es: enunciar preguntas-posibles respuestas (planteamiento de hipótesis)-resolución de las hipótesis (fase experimental u observacional)-conclusiones.

Para responder a las preguntas (2) de investigación planteadas y cumplir los objetivos del estudio, el investigador (a) debe seleccionar y desarrollar un diseño de investigación específico. Cuando se establecen y formulan hipótesis, los diseños sirven para someterlas a prueba.

El conducto en la investigación es: pregunta-objetivo- hipótesis-análisis. En los diagramas presentados en los libros de metodología de la investigación, se contextualiza la presencia de los estudios experimentales y no experimentales, en relación con el diseño del estudio.

La clasificación del tipo de diseño se hace según el manejo de las variables que determinan el alcance de la investigación (3) y la clasifican en experimental y no experimental. Significa con ello que en la investigación experimental se controlan algunas variables independientes para lo cual se diseña un experimento y los métodos son: diseños completamente aleatorios, diseños por bloques, cuadrados latinos y diseños factoriales. Todos estos métodos forman parte de un gran tema que se llama el *análisis de varianza, ANOVA* que permite relacionar una variable cuantitativa (dependiente) y una variable cualitativa (independiente). En el diseño observacional o no experimental los fenómenos se observan tal y como suceden en su contexto natural, sin que pueda controlarse algún factor.

**La investigación** es un proceso que nos lleva al conocimiento de algo mediante la aplicación de métodos y técnicas que nos permiten aprehender los elementos de ese algo y sus interrelaciones para comprenderlo, describirlo, explicarlo, predecirlo y transformarlo.

Si el conocimiento genera poder y autoridad y, éste es necesario para transformar la sociedad, la investigación debe ser el punto de partida y de confluencia de esa acción transformadora.

La investigación conduce al investigador (a) a un estado superior de conocimiento del fenómeno en estudio; además de una actitud positiva hacia la investigación, se debe tener cierta madurez que la dan una edad mayor y la experiencia; mi práctica me ha enseñado que estudiantes menores de 20 años de edad difícilmente comprenden el proyecto de investigación científica como un todo y la forma como se relacionan sus partes.

Si bien no toda persona tiene aptitudes y actitudes para la investigación, el profesional, debe llegar a manejar los elementos básicos del proceso de investigación para su ejercicio profesional.

Todos los docentes de la Universidad de Antioquia, somos responsables de la formación y acompañamiento de los estudiantes en su proceso de formación; es mayor la responsabilidad, para los docentes - investigadores y para quienes se encuentren vinculados a los grupos de investigación porque son quienes en últimas, pueden articular la docencia y la investigación por medio de los trabajos de grado de los estudiantes.

Todos sabemos que la pregunta de investigación es la que define el método que se utilizará en el proceso investigativo. Si se busca comprender e interpretar un fenómeno, el método debe ser el cualitativo, en tanto; si se busca describir o explicar un fenómeno el método debe ser el cuantitativo.

La experiencia sin embargo enseña, un trato diferencial entre la mayoría de investigadores de la Universidad de Antioquia para quienes el método cuantitativo ha sido hegemónico y apenas se empieza a aceptar el enfoque cualitativo en las investigaciones. Pienso que cualquiera sea el método de una investigación, ésta debe asumirse con rigurosidad para que el proceso de investigación gane en calidad.

En la metodología de investigación con enfoque cuantitativo, el instrumento de recolección de la información es una encuesta que se construye a partir de un cuestionario que recoge todas las preguntas que sirven para la medición de las variables, las cuales a su vez, surgen de los objetivos específicos y en algunos casos del marco teórico. Todos los libros de estadística traen recomendaciones para construir cuestionarios de preguntas que respondan a los objetivos porque así se gana en la validez de constructo.

En la investigación sobre temas médicos o psicológicos, es común que se utilicen escalas elaboradas por expertos con las siguientes características (2):

- a) La confiabilidad, se refiere al grado en que el instrumento en su aplicación repetida en el mismo individuo u objeto produce resultados iguales. Generalmente, se miden de dos formas: mediante el test-retest (coeficiente de correlación) con el cual se mide la estabilidad de la escala en sus puntajes individuales, a través del tiempo y con la consistencia interna (coeficiente Alpha de Cronbach) que mide la precisión con la que los ítems de la escala miden el constructo en estudio.
- b) La validez, se refiere al grado en que el instrumento mide la variable que pretende medir.
  - Validez de contenido: se refiere al grado en que un instrumento refleja un dominio específico de contenido.
  - Validez de criterio: se compara con algún criterio externo que mide lo mismo. Si el criterio se fija en el mismo momento, se habla de *validez concurrente*.
  - Validez de constructo: debe explicar el modelo teórico empírico, que subyace a la variable de interés.
  - Validez de expertos: los expertos definen si el instrumento mide la variable en cuestión.
- c) La validez total se gana con una mayor validez de contenido, validez de constructo y validez de criterio en el instrumento de medición.
- d) La relación entre la confiabilidad y la validez. El instrumento puede ser confiable pero no válido. Se requiere que se cumplan ambas características.
- e) Factores que pueden afectar la confiabilidad y la validez: la improvisación, utilizar instrumentos desarrollados en el extranjero que no han sido validados en el país de aplicación, es inadecuado para las personas a quienes se le aplica, las condiciones desfavorables en las que se aplica.

En lo posible, se deben utilizar instrumentos de los cuales se sepa que miden lo que pretenden medir, que son estables en el tiempo de aplicación; es decir que obtienen la misma información en intervalos de hasta 15 días de aplicación, que son contruidos por expertos en el tema y que han superado cualquier improvisación en la forma como se ordenan y relacionan las preguntas y el puntaje total que generan. No es recomendable construir un cuestionario de preguntas para indagar temas que tienen que ver con el comportamiento humano. En nuestra universidad, se impone cada vez, el utilizar escalas de medida que sean validadas, nacional e internacionalmente.

El proceso de búsqueda de escalas de medida para realizar una investigación puede tomar varios meses y muchas de ellas deben ser compradas, lo cual demanda que se presupueste en el costo total del proyecto. Generalmente, se consiguen escalas validadas internacionalmente y debe realizarse la validación a nivel nacional para que se garantice que la escala mida lo mismo que en cualquier otra parte.

En la presente ponencia se propone una metodología para la validación de escalas que han sido construidas en el extranjero y algunas de ellas han sido traducidas al idioma castellano.

### **Métodos**

Se revisan algunos artículos publicados en revistas indexadas sobre características psicométricas de los instrumentos o procesos de validación de escalas seguidos con el fin de determinar la ruta común para la validación de una escala. Se revisa una tesis de pregrado sobre la validación de una escala construida en el extranjero y traducida al castellano en Colombia. Se hace una propuesta metodológica para la validación de una escala.

## Resultados

### Revisiones de otras validaciones

- En un estudio en Chile (4), se tuvo una muestra por conveniencia con un tamaño,  $n=264$  de personas con edades entre 16 y 33 años, 57 % mujeres y 43 % hombres. La aplicación fue grupal en el horario de clase. Se utilizó consentimiento informado para los participantes en el estudio. Las escalas se aplican al mismo tiempo. Primero, se aplicó un Análisis de Componentes Principales, ACP, con rotación varimax para determinar tres factores. Luego, se realizó un Análisis Factorial Confirmatorio para verificar esa estructura factorial de la escala; se evaluó la consistencia interna de las escalas mediante el coeficiente Alpha de Cronbach. Se obtuvo la correlación ítem – test en la cual la distribución de cada ítem se correlaciona con la distribución del puntaje total de la escala. Se corre un modelo de regresión lineal múltiple para cada una de las escalas, orientación a la dominancia social (SDO) y la escala de nacionalismo.
- En un estudio con universitario en México (5), se obtuvo una muestra no probabilística de  $n=122$  estudiantes entre 17 y 25 años que solicitaron ayuda psicológica y que no estaban en tratamiento farmacológico. Se realizaron análisis factoriales para determinar la estructura factorial de la escala. Se calcularon coeficientes de correlación entre los ítems y el test y se obtuvieron medidas de consistencia interna de la escala y de las subescalas.
- En un estudio realizado en España (6), se obtuvo una muestra de  $n =584$  individuos; se realizó un Análisis Factorial Confirmatorio para verificar la estructura interna de la escala, luego se obtuvo el coeficiente Alpha de Cronbach para verificar la consistencia interna y por último se realizó un ANOVA entre el puntaje total y el sexo para verificar si hay diferencia en el puntaje total para los sexos.
- En un estudio con adolescentes españoles (7), se obtuvo una muestra de  $n=525$ , se aplicó un análisis factorial para determinar la estructura factorial y luego con ella, se aplicó un Análisis Factorial Confirmatorio, AFC para verificar que efectivamente esa estructura factorial es apropiada. Luego, realizó un ANOVA entre el puntaje total y el sexo para verificar si hay diferencia en el puntaje total para los sexos también se realizó lo mismo para los grupos de edad.
- En un estudio sobre uso problemático de las nuevas tecnologías en España (8), se obtuvo una muestra de  $n=2747$  individuos. Se realizó un análisis factorial con rotación varimax para determinar una estructura factorial; se obtuvo el KMO y se realizó la prueba de esfericidad de Bartlett para verificar la igualdad de varianzas. Se verificó la consistencia interna con el Alpha de Cronbach para la escala y para la escala sin el ítem. Se analizó la validez convergente mediante la correlación de la distribución de cada ítem con la subescala sin el ítem. Se realizó una validez clínica. , realizó un ANOVA entre el puntaje total y el sexo para verificar si hay diferencia en el puntaje total para los sexos.
- En un estudio con pacientes psiquiátricos hospitalizados realizado en Colombia (9), se obtuvo una muestra de tamaño  $n=149$ . Se realizó una prueba – preprueba con una muestra de  $n=30$  y un análisis interevaluador. La validez concurrente se obtuvo mediante la correlación entre el puntaje total de la escala y el de la escala ICG. La sensibilidad al cambio se verificó con la correlación entre el puntaje total de la prueba y el de la preprueba obtenidos en dos momentos clínicamente diferentes.
- En un estudio realizado con adolescentes escolarizados en San Juan de Pasto, Colombia (10). Se obtuvo una muestra aleatoria de 306 adolescentes escolarizados en las instituciones públicas y privadas para la aplicación final para el análisis de la escala fases 4 a7. El proceso de validación se ajustó a las siguientes fases: 1) fase 1: consolidación de la escala traducida al español. Traducción-retraducción; 2) fase 2: consolidación de la versión preliminar de la escala. Se revisan los ítems para que los valoren con respecto al tema, como: muy importante, importancia intermedia, menos importante; fase 3: pilotaje, utilizando una prueba piloto se analiza la comprensión de las instrucciones y de los ítems; fase 4: aplicación final; fase 5: validez del RFL-A adaptado al español mediante un Análisis Factorial exploratorio con rotación varimax para determinar una estructura factorial; se obtuvo el KMO. Para la validez convergente, se obtuvo el coeficiente de correlación con otra escala. Fase 6: fiabilidad del RFL-A: alpha de Cronbach, correlaciones ítem-prueba, alpha de Cronbach de la escala sin el ítem., alpha de Cronbach de las subescalas; fase 7: realizó un ANOVA entre el puntaje total y el sexo para verificar si hay diferencia en el puntaje total para los sexos.
- En un artículo sobre validación de escalas en salud (11) se propone que el proceso de validación debe tener las siguientes etapas: selección de la escala, traducción, pruebas preliminares para realizar ajustes (de ítems y utilidad), pruebas de su validez (validez de apariencia, validez de contenido, validez de criterio o concurrente, sensibilidad al cambio), pruebas de confiabilidad, determinación de su utilidad.

- En un artículo normas para el desarrollo y revisión de estudios instrumentales (12) se propone con una muestra de entre 5 y 10 individuos por ítem; debe ser mayor a 200 individuos y llevar el siguiente orden para presentar los resultados de la validación: validez de constructo (Análisis factorial confirmatorio), Fiabilidad (alfa de Cronbach y correlación ítem-subconstructo), confiabilidad test-retest.

### **Análisis previo al proceso de validación**

El primer paso es encontrar una escala que mida el fenómeno de nuestro de interés que haya sido construida sobre una base científica y que, ojalá, haya tenido una validación en Colombia. Si ese es el caso, podemos realizar nuestro estudio con absoluta seguridad de que nuestros resultados pueden publicarse sin temores en cualquier revista indexada del mundo.

Lo más común es que la escala construida que encontramos, está en un idioma diferente al castellano y ha sido validada en un país del extranjero. En este caso, se debe realizar una traducción de la escala, someterla a pruebas de expertos, validarla culturalmente, antes de ser aplicada a la población de estudio. La fase 1 de traducción – retraducción utilizada con adolescentes escolarizados en San Juan de Pasto (10) puede ser parte de la solución. Con una prueba piloto en la población de estudio, se puede analizar la comprensión de los ítems por parte de la población de estudio para validar culturalmente la escala.

Si la escala ha sido construida en un país del exterior de habla hispana y con una cultura par a la colombiana, la validación cultural no es necesaria y se debe proceder a la validación en el país de aplicación.

### **Propuesta metodológica para validar una escala**

Se parte de un proyecto de investigación que cuenta con el aval de algún comité de ética y que cuenta con una escala en idioma castellano que debe ser validada para Colombia. Se analiza la forma como está construida, sus subconstructos o subescalas, los valores de los coeficientes de la primera validación y la población en la cual puede ser aplicada la escala.

**El primer paso** es diseñar una muestra aleatoria para estimar la proporción de interés si se mide en un nivel, que sea mayor a 200 personas ( $n > 200$ ); es decir, el error permisible máximo en la estimación de la proporción es del 7 % y en forma equivalente para la estimación del puntaje promedio de la escala. Algunos autores sugieren como criterio que tenga entre 5 y 10 individuos por cada ítem de la escala) (12). Con esta muestra, se procede luego a verificar la estructura factorial mediante un Análisis Factorial Confirmatorio, AFC, si se conocen los subconstructos. Si no hay subconstructos de manera explícita, se debe realizar previamente un Análisis Factorial para determinar la estructura factorial subyacente y luego el Análisis Factorial Confirmatorio.

**En segundo lugar**, Para verificar la fiabilidad de la escala, se verifica su consistencia interna mediante el alfa de Cronbach para todos los ítems. También se calcula el coeficiente de correlación entre cada ítem y el puntaje total del subconstructo.

**En tercer lugar**, se determina una escala par que mida lo mismo en la población de interés, y que haya sido validada en Colombia.

**En cuarto lugar**, se diseña una muestra aleatoria de tamaño mayor o igual a 50 para que garantice que se conserva la estructura factorial de la escala (13). Con esta muestra se verificará la estabilidad de la escala en el tiempo y la validez concurrente correlacionando los puntajes totales de las dos escalas.

Luego de obtener el consentimiento informado de los individuos del estudio, se aplican las dos escalas simultáneamente para realizar la validación concurrente con la muestra aleatoria de tamaño mayor o igual 50. Se aplican la escala de interés y la escala par en un solo cuestionario construido por el equipo de investigación en el cual se agregan variables como, sexo y edad. Se obtiene así, el pretest o la preprueba.

**A los cuatro días** de la primera aplicación, se aplica al mismo grupo de personas la escala de interés y se obtiene así el retest o la prueba.

Para verificar la normalidad de los puntajes de cada uno de los instrumentos, se realiza la prueba de normalidad de Kolmogorov – Smirnov con corrección de Liliefors en el software estadístico SPSS (14). Esta prueba se hace también para la distribución de cada uno de los ítems de la escala de interés.

**En quinto lugar**, verificar la reproducibilidad test- retest mediante el coeficiente de correlación entre los puntajes totales de la escala de interés, obtenidos en las dos aplicaciones con cuatro días de intervalo.

**En sexto lugar**, verificar la validez de convergencia entre las dos escalas mediante el coeficiente de correlación de Spearman, donde el contraste de hipótesis es:  $H_0$ : no existe correlación lineal entre los ítems.  $H_1$ : existe correlación lineal entre los ítems. Por lo tanto si el valor p es menor que el 5 %, la correlación es diferente de cero. Además, se obtiene el índice de Kappa para determinar la concordancia entre los niveles de riesgo de suicidio de las dos escalas. Todo esto se hace en el SPSS también (14).

**Calcular el índice de concordancia de Kappa** desde una tabla de contingencia 2x2 en la cual se colocan en las filas los valores que representan la existencia del evento en la escala de interés y en las columnas los valores que representan la existencia del evento en la otra escala. El contraste de hipótesis para el índice de kappa es:  $H_0$ : no existe acuerdo entre los puntajes de ambas escalas.

$H_1$ : existe acuerdo entre los puntajes de ambas escalas.

**En séptimo lugar**, determinar la utilidad de la escala desde su valoración y desde su aplicación.

### **Bibliografía**

1. Álvarez R. Ensayos clínicos. Diseño, análisis e interpretación. Editorial Díaz de Santos. España; 2005.
2. Hernández R, Fernández C, Baptista P. Metodología de la investigación. 5ª edición. Mc Graw-Hill. Perú; 2010.
3. Medina María. La investigación aplicada a proyectos. Identificación del proyecto y formulación de la investigación. 2ª edición. Centro de investigaciones para el desarrollo CAYÉ. Volumen I. Bogotá; 2014.
4. Cárdenas M. Adaptación y validación de la versión abreviada de la escala de autoritarismos de derechos (RWA) e una muestra chilena. Revista de psicología de la Universidad de Chile. XIX (1): 61-79; 2010.
5. González S, Díaz A, González C, González J. Características psicométricas de la escala de ideación suicida de Beck (ISB) en estudiantes universitarios de la ciudad de México. Salud Mental. 23 (2): 21-30. Instituto Nacional de Psiquiatría de la Fuente Nuñez. México.
6. Sicilia A, Alías A, Ferriz R, Moreno J. Spanish adaptation and validation of the exercise addiction inventory (EAI). Psicotema. 25 (3): 377-383. Spain; 2013.
7. Casas J, Ruiz R, Ortega R. Validation of the internet and social networking experiences questionnaire in spanish adolescent. International journal of clinical and healthpsychology. 13: 40-48. Córdoba, Spain; 2013.
8. Labrador F, Villadangos S, Crespo M, Becoña E. Desarrollo y validación de uso problemático de nuevas tecnologías. Anales de psicología. 29 (3): 836-847. España; 2013.
9. Sánchez R, Ibañez M, Pinzón A. Análisis factorial y validación de la versión en español de la escala briefpsychiatric rating scale in Colombia. Biomédica. Instituto Nacional de Salud. Bogotá, Colombia. 25: 120-128; 2005.
10. Moreno A. Propiedades psicométricas del inventario razones para vivir (RFL-A, Osman et al., 1998), adaptado al español en una muestra de estudiantes de bachillerato de la ciudad de San Juan de Pasto. [Tesis de pregrado]. Universidad de Nariño. Facultad de Ciencias. Programa de Psicología. San Juan de Pasto, Colombia; 2008.
11. Sánchez R, Echeverry J. Validación de escalas de medición en salud. Rev. Salud pública; 2004; 6 (3):302-3018.
12. Carretero H, Pérez C. Normas para el desarrollo y revisión de estudios instrumentales. International journal of clinical and healthpsychology; 2005; 5(3):521-551.
13. Hair J, Anderson R, tatham R, Black W. Análisis multivariante. 5ª edición. Pearson-Prentice Hall. Madrid; 1999. p. 100.
14. Pardo A, Ruiz M. Análisis de datos con SPSS 13 base. España: Mac Graw – Hill; 2005.