

# APUNTES SOBRE METODOLOGÍA DE LA INVESTIGACIÓN CONFIABILIDAD?

*Arturo Barraza Macías*

Profesor del área de postgrado de la Universidad Pedagógica de Durango; actualmente se desempeña como Coordinador del Programa de Investigación.

Bajo la denominación genérica de confiabilidad se agrupan todo un conjunto de métodos y técnicas utilizadas por los investigadores para estimar el grado de precisión con el que están midiendo sus variables; “la confiabilidad indica la consistencia del proceso de medición o de los resultados. Por lo tanto, no es correcto hablar de la confiabilidad de la una prueba o instrumento” (Bonilla, 2006; p. 1).

La consistencia, precisión o estabilidad temporal del proceso de medición suele expresarse mediante un coeficiente de confiabilidad que varía desde 0.00 a 1.00. Este coeficiente, gracias al apoyo de los programas computarizados de estadística, es relativamente fácil de obtener, sin embargo, una vez obtenido, el verdadero problema es, más allá de la obviedad de que un puntaje alto es mejor, ¿cómo valorar un coeficiente de confiabilidad?

La respuesta a esta pregunta constituye el eje central de este artículo que se divide en dos partes: en un primer momento ofrezco una serie de referentes conceptuales para situar la discusión y contextualizar adecuadamente la respuesta que se brinda en un segundo momento.

## Referentes Conceptuales

Hace más de cuarenta años, Robert L. Thorndike (en Muñiz, 1998) empezaba su famoso trabajo sobre confiabilidad con estas palabras: “Cuando medimos algo, bien sea en el campo de la física, de la biología o de las ciencias sociales, esa medición contiene una cierta cantidad de error aleatorio. La cantidad de error puede ser grande o pequeña, pero está siempre presente en cierto grado” (p. 6). Sus palabras siguen vigentes en la actualidad, pues en lo esencial los problemas de la medición cambian poco, aunque los instrumentos de medida vayan y vengan.

Cuando un investigador aplica un test, una escala, un inventario o cualquier otro instrumento de medida a una persona, obtiene una cierta puntuación, que por razones obvias se denomina puntuación empírica. ¿Cómo estar seguros de que esa puntuación obtenida es la que verdaderamente le corresponde a esa persona en esa prueba? En otras palabras, ¿cuánto error afecta a esa puntuación empírica?

Responder estas preguntas es el objetivo de la confiabilidad sin embargo para algunas personas tales interrogantes parecerían

incontestables, pues, al fin y al cabo, el error cometido, sea el que sea, está diluido en la puntuación empírica y no hay manera de separarlo. Sin embargo, se han creado un conjunto de estrategias con este propósito.

En el caso de las variables psicológicas, la propuesta pionera y más fructífera para la estimación de los errores fue hecha a principios de siglo por Spearman (en Muñiz, 1998 y 2003) y se denomina hoy Modelo Lineal Clásico, dando origen a todo un enfoque general sobre los tests que suele conocerse como Teoría Clásica de los Tests (Muñiz, 2003).

La propuesta de Spearman (en Muñiz, 1998) se basa, entre otros supuestos, en la definición del concepto de tests paralelos (aquellos que miden lo mismo aunque utilizando distintos ítems). A partir de esta propuesta surgen y se desarrollan una serie de estrategias para la medición de la confiabilidad entre las que destacan:

- La correlación entre dos formas paralelas del test,
- La correlación entre dos mitades aleatorias del test corregida mediante la fórmula de Spearman-Brown, y
- La correlación entre dos aplicaciones del mismo test a una muestra de personas.

De manera simultánea a estas estrategias se desarrollaron un conjunto de fórmulas para medir la consistencia interna del test, por lo que se estableció un nuevo rubro de la confiabilidad centrada en la consistencia interna. Entre las fórmulas más conocidas tenemos el coeficiente KR20 y KR21 (Kuder y Richardson), o el popular Coeficiente Alfa de Cronbach,

Aunque existen distintas formas de evaluar la confiabilidad de la medición realizada por el instrumento, se puede afirmar que las estrategias test-retest y las de consistencia interna suelen ser las más utilizadas (Muñiz, 1996 y 2003).

La estrategia test-retest consiste en la aplicación de un mismo instrumento a una misma muestra de sujetos en al menos dos momentos diferentes. No existe un criterio único respecto de cual debe ser el lapso adecuado entre la primera y segunda aplicación. Este se establece en función de los fines del instrumento o la investigación (Muñiz, 1996).

Por su parte, las estrategias de consistencia interna permiten aplicar el instrumento tan sólo una vez. Las dos estrategias más utilizadas son la confiabilidad por mitades y el coeficiente alfa de cronbach.

#### a) Confiabilidad por mitades.

Una forma de evaluar la consistencia interna de un instrumento es mediante el procedimiento de dos mitades; en este caso todos los participantes responden una sola vez a la prueba. Sin embargo, se le asignan dos puntuaciones a cada participante. Para lograr esto, la prueba se divide en dos mitades. Las dos puntuaciones se pueden calcular dividiendo la prueba a la mitad dependiendo del número de ítems o se puede dividir tomando los ítems pares e impares. Se obtiene la correlación entre las dos puntuaciones resultantes y se utiliza la fórmula de Spearman-Brown para estimar la confiabilidad del instrumento completo.

Este tipo de confiabilidad es conocido como confiabilidad por mitades (Cozby, 2005), método de las mitades o splitt-half method (Briones, 2001),

método de mitades partidas o splitt-halves (Hernández, Fernández y Baptista, 2006) o simplemente Split-half (Bonilla, 2006)

b) Confiabilidad en Alfa de Cronbach

El procedimiento más utilizado (Muñiz, 1996 y 2003) para evaluar la consistencia interna de un test es el coeficiente Alfa de Cronbach.

La confiabilidad en Alfa de Cronbach es una generalización de las fórmulas KR para ítems de alternativas múltiples. El Alfa de Cronbach, cuyo cálculo emplea el promedio de todas las correlaciones existentes entre los ítems del instrumento que tributan al concepto latente que se pretende medir

### Valoración del coeficiente de confiabilidad

Una vez obtenido un coeficiente de confiabilidad, la pregunta obligada es ¿qué tan elevada debe ser la confiabilidad? La respuesta a esta interrogante no es sencilla y como afirma Hogan (2004) la única respuesta concluyente es "todo depende".

En concreto depende de lo que usted desee hacer con la prueba: es como preguntar "¿qué tan alta debe ser una escalera?"... todo depende: ¿necesita cambiar una bombilla que no está a su alcance o subir al techo de un edificio de tres pisos? (Pág. 112).

Si quien aplica una prueba pretende tomar decisiones sobre una persona (diagnóstico psicológico, selección de personal, licencia para ejercer una profesión, etc.) se requiere una prueba de alta confiabilidad (.90 como mínimo aceptable y .95 como la norma deseable); si en cambio el uso de la prueba es para la investigación se requiere una confiabilidad moderada (.80 se considera adecuada) (Nunnally y Bernstein, en Hogan, 2004).

Por su parte Rosenthal (en García, 2005) propone una confiabilidad mínima de .90 para tomar decisiones sobre la vida de las personas y una confiabilidad mínima de .50 para propósitos de investigación.

En esta misma línea de discusión, que tiene a los propósitos como referente para determinar el nivel de confiabilidad aceptable, Kaplan y Sacuzzo (en Hogan, 2004) señalan que la confiabilidad en el rango de .70 y .80 es lo suficientemente buena para cualquier propósito de investigación, mientras que la confiabilidad para una prueba que se utilice para tomar una decisión que influye sobre el futuro de una persona debe ser superior a .95.

Más allá de estas recomendaciones algunos autores nos ofrecen escalas completas para la valoración del coeficiente de confiabilidad.

De Vellis (en García, 2005), plantea la siguiente escala de valoración:

- Por debajo de .60 es inaceptable
- De .60 a .65 es indeseable.
- Entre .65 y .70 es mínimamente aceptable.
- De .70 a .80 es respetable.
- De .80 a .90 es muy buena

Por su parte Murphy y Davishofer (en Hogan, 2004) ofrecen la siguiente escala:

- Alrededor de .90 es un nivel elevado de confiabilidad.

- La confiabilidad de .80 o superior puede ser considerada como moderada.
- Alrededor de .70 se considera baja.
- Inferior a .60 indica una confiabilidad inaceptablemente baja.

#### A manera de cierre

Las recomendaciones sobre el nivel de confiabilidad adecuado, en función del propósito de la prueba, y las escalas propuestas, nos ofrecen una respuesta concreta a la pregunta que originó este trabajo, sin embargo, algunos lectores seguirán preguntándose ¿cuál es la respuesta correcta?

Dejo a mis lectores la oportunidad de decidir cual de las posibles respuestas consideran más adecuada, yo por mi parte, me adhiero a la postura de Thorndike y Hagen (1978), quienes en su clásico libro "Tests y técnicas de medición en psicología y educación" consideran que la valoración de la confiabilidad de un proceso de medición debe de hacerse siempre en función de la superioridad que manifieste sobre la confiabilidad de otros procesos de medición existentes con relación a la misma variable.

En otras palabras, la confiabilidad obtenida debe ser mayor a la que se ha obtenido en otros procesos de medición que utilizan otros instrumentos para medir la misma variable.

#### Lista de Referencias

- Bonilla Rodríguez Víctor E. (2006), *Confiabilidad*, en el boletín informativo INEVA en acción, Vol. 2, No. 3, pp. 1-3.
- Briones Guillermo (2001), *Métodos y técnicas de investigación para las ciencias sociales*, México, Trillas.
- Cozby Paul C. (2005), *Métodos de investigación del comportamiento*, México, Mc Graw Hill.
- García Cadena Cirilo H. (2006), "La medición en ciencias sociales y en la psicología", en *Estadística con SPSS y metodología de la investigación*, de René Landeros Hernández y Mónica T. González Ramírez (comp.), México, Trillas.
- Hernández Sampieri Roberto, Fernández Collado Carlos y Baptista Lucio Pilar (2006), *Metodología de la Investigación*, México, Mc Graw Hill.
- Hogan Thomas P. (2004), *Pruebas psicológicas*, México, El Manual Moderno.
- Muñiz José (1996). *Psicometría*. Madrid, España, Editorial Universitaria
- Muñiz José (1998), *La medición de lo psicológico*, en la revista Psicothema, Vol 10, No. 1, pp.1-21
- Muñiz José (2003), *Teoría Clásica de los Test*, Madrid, España, Ediciones Pirámide.
- Thorndike Robert L. y Hagen Elizabeth (1978), *Tests y técnicas de medición en psicología y educación*, México, trillas.



Ilustración 1 La interpretación de los fenómenos J. M. L. A.