Inv Ed Med 2013;**2(6)**:107-111

Investigación en Educación Médica

www.elsevier.com.mx

METODOLOGÍA DE INVESTIGACIÓN EN EDUCACIÓN MÉDICA

Confiabilidad en la medición

Lucy María Reidl-Martínez.

División de Estudios de Posgrado, Facultad de Psicología, Universidad Nacional Autónoma de México. México D.F., México.

Recepción 25 de diciembre de 2012; aceptación 25 de enero de 2013

PALABRAS CLAVE
Confiabilidad; investigación educativa; análisis de datos; México.

Resumen
En la investigación educativa de tipo cuantitativo, los instrumentos para la recopilación de información deben llevar a la validez y confiabilidad de los resultados. Una pregunta que los investigadores se hacen con frecuencia es la referida, a qué tan alto o de qué tamaño tiene que ser el coeficiente de confiabilidad. Esto dependerá del propósito e importancia de las decisiones que se tomarán con base en los puntajes obtenidos por los sujetos. La confiabilidad es un atributo necesario en todas las pruebas cuantitativas que se utilicen para evaluar a alguien. El presente escrito tiene como propósito abordar la confiabilidad de la medición. Para lograrlo se revisan los siguientes puntos: teoría de la medición del error, modelo dominio-muestra, estabilidad temporal o coeficiente de estabilidad, formas paralelas o coeficiente de estabilidad y equivalencia y los coeficientes de consistencia interna más usuales.

KEYWORDS
Reliability; educational research; data analysis; Mexico.

Reliability of measurement

Abstract
In quantitative educational research, instruments for data collection should lead to the validity and reliability of results. A question that researchers are frequently referred to is how high or how large must the reliability coefficient be. This depends on the purpose and importance of the decisions to be made, based on the scores obtained by the subjects. Reliability is a necessary attribute in all quantitative tests used to assess someone. The present paper aims to address the reliability of the measurement. To achieve the following chapters are reviewed: theory of measurement error, domain-sample model, temporal stability or stability coefficient, parallel forms or stability coefficient and equivalence and some of the most used internal consistency coefficients.

Correspondencia: Lucy María Reidl-Martínez. División de Estudios de Posgrado, Facultad de Psicología, Universidad Nacional Autónoma de México. Edificio de los Consejos Académicos, Ciudad Universitaria, Delegación Coyoacán, C.P. 04510, México D.F., México. Correo electrónico: lucym@unam.mx

ISSN: 2007-5057 - see front matter © 2013 Facultad de Medicina Universidad Nacional Autónoma de México. Publicado por Elsevier México. Todos los derechos reservados.

Un elemento esencial en los proyectos de investigación educativa son los instrumentos para la recopilación de información. Estos deben llevar a la validez y confiabilidad de los resultados. El presente documento tiene como propósito abordar la confiabilidad de la medición.

Teoría de la medición del error

Es en el campo de la psicología donde se desarrolló desde hace ya tiempo, alrededor de los años 50 del siglo pasado, *la teoría de la medición del error*, con el objetivo de poder construir instrumentos que midieran de manera confiable y válida, muchas de las variables que estudia esta disciplina: la inteligencia, la personalidad, las actitudes, las opiniones, la ansiedad, los diferentes motivos que impulsan al ser humano, las expresiones emocionales, los sentimientos, los síntomas y los trastornos de personalidad, las habilidades, las intenciones, las creencias, entre otras muchas cuestiones de interés para la disciplina.[1]

Y un tiempo después, se desarrolla otro modelo de medición, que al mismo tiempo que da mayor sentido al primero, resuelve las deficiencias que éste tenía, surgiendo así, el *modelo dominio-muestra*. Al ser la psicología una disciplina que pretende explicar el comportamiento humano, así como sus sentimientos, estados de ánimo, sus creencias y expectativas, y muchas cuestiones más, también se ha visto utilizada o abordada por otras disciplinas, como es el caso de la medicina humana, la sociología, la mercadotecnia, la política, entre otras.

Lo anterior llevó a definir a la medición de diversas maneras: a) como la asignación de números a las cantidades de propiedades de los objetos de acuerdo con reglas dadas cuya validez puede probarse empíricamente; b) en términos más sencillos se puede definir como la asignación de la magnitud en que un objeto o sujeto posee cierta propiedad, con ayuda del sistema numérico[2] o, c) en términos de dos conceptos relativamente sencillos: de un conjunto de reglas para asignar símbolos a objetos de manera que representen numéricamente, cantidades de atributos (escalamiento) y definir si los objetos pertenecen a la misma o diferentes categorías, en relación con un atributo dado (clasificación).[3]

En medición se parte del supuesto de que el procedimiento de aplicar pruebas objetivas, consiste en presentar a la persona un número de reactivos (preguntas o aseveraciones) que debe contestar, y cada respuesta se califica como correcta/incorrecta, de acuerdo/en desacuerdo (1 y 0), o en una escala que puede ir del 1 al 5 (7, 9 o 4, 6, 8 o algún otro número), donde en la medida en que incrementa el valor numérico escogido por el respondiente, se reporta una mayor cantidad de la variable (cualquiera que ésta sea) poseída o descriptora del sujeto.[4]

Los números usados pueden proporcionar diferentes cantidades de información, que permite distinguir entre tres niveles de medida; las magnitudes o números asignados pueden corresponder a una escala ordinal, una de intervalo o una de proporción. La primera proporciona información sobre el orden de los objetos o sujetos respecto al rasgo o variable que se mide; la segunda proporciona información acerca del tamaño de las diferencias entre los objetos o sujetos respecto a la magnitud del rasgo o variable medida; y la tercera da información no sólo del orden de rango de los sujetos u objetos y del tamaño relativo de las diferencias, sino también de la relación entre las proporciones o mediciones mismas.

La calificación obtenida por el respondiente será la suma de los números escogidos por él, a lo largo de los diversos reactivos, preguntas, aseveraciones o problemas respondidos. Y es aquí donde surge el problema de determinar la exactitud de dicha medición. El error de medición en este tipo de instrumentos, como se puede imaginar, es mucho mayor que lo que podría ser en otras áreas como la física.

Para atender esta problemática se requiere hacer una serie de suposiciones al respecto de la relación existente entre los *puntajes verdaderos* (T) y los de *error* (E). Se parte del supuesto de que cada calificación observada (X) tiene dos componentes: a) la calificación verdadera del sujeto en la variable que se está midiendo (T), que es más o menos estable siempre y cuando se esté midiendo lo mismo en diversas ocasiones; y b) el componente de error (E), que puede deberse a que el sujeto responda de manera correcta el reactivo o pregunta por cuestiones de azar, o que responda correctamente por que conoce la respuesta (pruebas de conocimiento); en el caso de medir por ejemplo, actitudes o rasgos de personalidad, el sujeto puede responder, por ejemplo, de manera socialmente deseable (atribuyéndose una actitud o rasgo valorado como negativo con menor intensidad a lo que realmente siente o lo describe), o independientemente de ello. Cualquiera que sea la situación, se supone que las calificaciones o puntajes obtenidos por los individuos (puntajes observados, X), quedan constituidos por la suma de los puntajes verdaderos (T) y los de error (E). Existen dos tipos de error, el aleatorio y el sistemático o constante. Si el procedimiento de medición da puntajes consistentemente mayores (o menores) de lo que deberían de ser, se denomina error constante; si en ocasiones el error es grande, en otras es pequeño, o a veces es positivo y otras es negativo, se habla del error aleatorio o al azar.

Los supuestos básicos de la teoría de la medición por medio de pruebas o escalas, parten del supuesto del comportamiento aleatorio del error, que señala que si éste se presenta en un número suficientemente grande de casos, en la medida en que este número aumente, la media o promedio de la suma de estos errores se acerca a cero. Este supuesto puede referirse a reactivos o preguntas, o a personas o sujetos que los contesten. Derivado de lo anterior, también se puede suponer que en la medida en que aumente el número de casos, la correlación entre los puntajes verdaderos y los de error, se acerca a cero.

Otra manera de definir el error aleatorio tiene que ver con la correlación del error en una prueba y el error de otra prueba (paralela, que mida lo mismo). Es decir, el modelo plantea que la correlación entre dos conjuntos de errores aleatorios es cero, o se acerca a cero en la medida en que incrementa el número de casos. Aunque este supuesto sólo es cierto en la medida en que el número de casos se acerca al infinito, en la práctica, se supone que esto se mantiene para cualquier conjunto dado de datos.

Resumiendo, los supuestos de los que se parte en la teoría de la medición del error aleatorio son los siguientes:

a) el error promedio es igual a cero; b) la correlación entre la calificación del error y la calificación verdadera es cero; y c) la correlación entre los errores de una prueba y los de otra prueba paralela es igual a cero. Obviamente estos supuestos sólo se mantienen cuando el número de casos es muy grande (cercano al ∞), pero se utilizan como punto de partida para cualquier conjunto de datos obtenido por pruebas, escalas o cuestionarios.

Entre las diversas teorías de la medición del error, algunas suponen que la calificación verdadera obtenida por el sujeto sería la calificación promedio que se obtendría si se repitiera la aplicación de la prueba en un número infinitamente grande de ocasiones, situación que nunca ocurre. Este error en la medición se denomina error estándar de la medición, de acuerdo a algunos modelos de medición, y se considera que se distribuye de manera normal alrededor de las calificaciones verdaderas, y que es una constante para cualquier objeto que se mida.[3,5]

Modelo dominio-muestra

Otro modelo particular que da lugar a poder hablar de calificaciones verdaderas es el que parte del muestreo de un dominio donde se construye una prueba seleccionando un número específico de preguntas al azar, de un conjunto infinitamente grande y homogéneo de reactivos. Este modelo, denominado modelo dominio-muestra, parte del supuesto de que la variedad de reactivos o preguntas que componen a una prueba, tienen efectos semejantes a aquellos que procedieran de un muestreo realmente aleatorio de los mismos. Señala también, que el propósito de cualquier medición es el de estimar la medida que se obtendría si uno utilizara todos los reactivos del dominio. La calificación verdadera correspondería a la que obtendría el sujeto si respondiera a todas las preguntas del dominio, y por ello una prueba, (un conjunto de reactivos o preguntas) será confiable en el grado en que la calificación que arroje correlacione en buena medida con las calificaciones verdaderas.

Como modelo que es, supone que existe una matriz de correlaciones infinitamente grande, y la correlación promedio de esta matriz indica la medida en que existe un núcleo o elemento entre todas las variables (preguntas o reactivos); y la dispersión de los valores de las posibles correlaciones, señala el grado en que las variables varían en compartir este núcleo o elemento común.[3,5,6]

El modelo de pruebas paralelas es una de las alternativas del modelo dominio-muestra, y supone que dos o más pruebas producen calificaciones verdaderas iguales, pero que generan errores de medición independientes para cada una de ellas. El uso de formas paralelas o alternas de una prueba es una manera de evitar las dificultades de la confiabilidad de *test-retest*, que queda establecida por la correlación entre ambas aplicaciones, donde la varianza de error puede provenir de fluctuaciones aleatorias de la ejecución de una sesión a otra, de condiciones no controladas de la aplicación, condiciones de los examinados (fatiga, tensión, etc.).[6]

Confiabilidad

Ahora bien, de los supuestos anteriores derivan las diversas formas de determinar la confiabilidad de las pruebas. La confiabilidad de una prueba se refiere a la consistencia de las calificaciones obtenidas por las mismas personas en ocasiones diferentes o con diferentes conjuntos de reactivos equivalentes.[7] El concepto de confiabilidad subyace al error de medición de una sola calificación que permite predecir el rango de fluctuación que puede ocurrir en la calificación de un sujeto, como resultado de factores irrelevantes aleatorios, como ya se ha mencionado.

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. Esencialmente, cualquier condición que es irrelevante al propósito de la prueba representa error de la varianza; cuando el investigador trata de mantener condiciones de prueba uniformes, controlando el ambiente en el que se lleva a cabo, las instrucciones, los tiempos límites, el "*rapport*" y otros factores similares, está tratando de reducir el error de la varianza y hacer que las calificaciones de las pruebas sean más confiables. Como esto es imposible de conseguir aunque se tuvieran las condiciones óptimas, ninguna prueba es totalmente confiable y por ello, cada una de ellas debe establecer su confiabilidad. Esta medida de confiabilidad es característica de la prueba si se aplica en condiciones estándar, y en sujetos similares a aquellos con los que se estableció la muestra normativa. Por ello, se deben especificar las características de la tal muestra junto con el tipo de confiabilidad que se estableció en cada ocasión en que se construye o adapta una prueba para una muestra con características diferentes a las de la muestra original.

Existen diferentes tipos de confiabilidad: a) estabilidad temporal o coeficiente de estabilidad; b) formas paralelas o coeficiente de estabilidad y equivalencia; c) división por mitades o coeficiente de consistencia interna; y d) consistencia interna pura.[1-9]

a. Estabilidad temporal o coeficiente de estabilidad

Cada una de ellas se calcula teniendo un objetivo en mente. Por ejemplo, la estabilidad temporal indica el grado en el que las calificaciones de una prueba se ven modificadas por fluctuaciones aleatorias diarias en la condición del sujeto o en el ambiente de prueba. Esta estabilidad depende en parte de la longitud del intervalo en el que se mantiene, y es indispensable establecerla, si el objetivo del investigador es medir cambios a lo largo del tiempo. Es decir, asegurar que si se presentan cambios en la variable de interés, se debieron al paso del tiempo (por ejemplo, la hora del día o debido al desarrollo) y no al instrumento de medición. En este caso, los mismos sujetos responden a dos administraciones diferentes de la misma prueba, y se espera que la variable no cambie con el transcurso del tiempo, la correlación entre los puntajes obtenidos tendrá que ser alta.[5]

b. Formas paralelas o coeficiente de estabilidad y equivalencia

Las formas paralelas o equivalentes, representan otro tipo de confiabilidad que se requiere cuando se espera

que una situación (experimental o cotidiana), modifique la variable de interés, en un lapso muy corto, que no permitiría aplicar el mismo instrumento, pues los sujetos podrían recordar las respuestas dadas con anterioridad y/o contestar diferente por creer que es lo que se espera de ellos, o contestar de manera muy semejante a como lo hicieron con anterioridad, porque recuerdan las respuestas dadas en la primera ocasión. En este caso, se necesitan dos versiones del instrumento, que midan lo mismo, pero con diferentes reactivos, estímulos o preguntas. Al coeficiente que se calcula para determinar la medida en que se mide los mismo con ambas versiones, se denomina coeficiente de equivalencia.[5]

c. División por mitades o coeficiente de consistencia interna

La confiabilidad de división por mitades, se determina dividiendo a la prueba en mitades, asegurando que los reactivos o preguntas se hayan ordenado de acuerdo a su grado de dificultad (de los más fáciles a los más difíciles); se constituye una especie de prueba paralela, con los reactivos pares en uno de los conjuntos, y los impares en el otro, asegurando de alguna manera que los reactivos sean igualmente difíciles en ambos conjuntos, o en términos estadísticos, propiciando que las distribuciones de ambos conjuntos tengan medias y varianzas semejantes. El coeficiente de consistencia interna se determina en este caso con la fórmula de Spearman-Brown, que sólo se puede aplicar a pruebas homogéneas y sin límite de tiempo para resolverlas.[2,4,5,9]

Si la prueba es heterogénea, el coeficiente de equivalencia debe calcularse con las mitades de la prueba, igualada, no sólo en términos de dificultad sino también en términos del contenido.

En las pruebas con límite de tiempo, los individuos necesariamente no tienen el tiempo suficiente para resolver todos los reactivos (pruebas de velocidad *vs.* las pruebas de poder, en las que no existe tiempo límite) y por ello la determinación de la confiabilidad quedará establecida por la correlación entre los reactivos pares y nones.[2]

La consistencia interna tiene que ver con la equivalencia de los reactivos y la homogeneidad de los mismos; y está basada en la consistencia de las respuestas de los sujetos en todos los reactivos de la prueba. Un coeficiente de consistencia interna proporciona tanto una medición de equivalencia como de homogeneidad. De tal manera que dos pruebas que tienen una alta confiabilidad en términos de formas paralelas o coeficientes por mitades, pueden variar en sus coeficientes de consistencia interna si difieren en el grado de homogeneidad de sus reactivos. De hecho, la diferencia de la consistencia interna entre mitades de la prueba y el coeficiente de consistencia interna, se debería utilizar como un indicador de la heterogeneidad de los reactivos o aseveraciones de una prueba o escala. Esto último sucede con mayor frecuencia en el caso de que una prueba mida diversos rasgos (de personalidad) o habilidades (de inteligencia), por ejemplo.

d. Consistencia interna pura

El procedimiento más común para determinar la consistencia interna de instrumentos o pruebas constituidas por respuestas dicotómicas (correcto-incorrecto; de acuerdo-en desacuerdo) es el desarrollado por Kudder y Richardson (KR-20), que se calcula a partir de una sola administración de una prueba. Esta técnica se basa en el examen de la ejecución en cada uno de los reactivos o preguntas de la prueba. A menos de que los reactivos sean muy homogéneos, este coeficiente siempre será menor que el de la confiabilidad por mitades. Ambos coeficientes (mitades y KR-20) igual que cualquier otro coeficiente de confiabilidad derivado de una sola administración de una sola forma de la prueba, se denominan coeficientes de consistencia interna. Sin embargo, la información que proporcionan los diferentes modelos de determinación de la consistencia interna, no es idéntica, por lo cual siempre se deberá de establecer cuál coeficiente se utilizó para determinarla.

Hasta ahora, se ha hablado de diferentes tipos de coeficientes de confiabilidad. De acuerdo al procedimiento seguido, cuando se repite una prueba, con la misma forma pero en diferente ocasión, se habla del coeficiente de estabilidad, y la varianza de error queda explicada por la fluctuación temporal; cuando se repite la prueba pero con una forma paralela, y en diferente ocasión, se habla de un coeficiente de estabilidad y de equivalencia, siendo la fuente de error la fluctuación temporal y la especificidad de los reactivos; cuando se repite la prueba, la misma forma y en la misma ocasión, se habla de un coeficiente de equivalencia y la fuente de error es la especificidad de los reactivos; cuando se divide a la mitad una prueba (pares-nones, u otra división paralela) se habla del coeficiente de consistencia interna y la fuente de error es la especificidad de los reactivos; y cuando se utilizan las pruebas Kudder-Richardson y el Alfa de Cronbach, se habla de un coeficiente de consistencia interna y la fuente del error son la especificidad de los reactivos, así como su heterogeneidad.[3,8,9]

Coeficiente Alfa de Cronbach

Ahora bien, el coeficiente Alfa, desarrollado por Cronbach en 1951, puede considerarse como equivalente a la media de todas las posibles correlaciones por mitades, corregidas con la fórmula de Spearman-Brown, y se utiliza en el caso de aquellas pruebas que tienen más de dos opciones de respuestas posibles. Su fórmula es:

$$r_\alpha = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum \sigma j^2}{\sigma^2}\right)$$

donde: r_α = coeficiente alfa.
k= número de reactivos.
σj^2= varianza de un reactivo.
$\sum \sigma j^2$ = suma de las varianzas de cada reactivo.
σ^2 = varianza de todas las calificaciones de la prueba

La prueba alfa es la estadística preferida para obtener una estimación de la confiabilidad de consistencia interna, y se usa como una medida de confiabilidad, en parte, debido a que se requiere de una sola aplicación al grupo de sujetos. Los valores típicos de esta prueba van de 0 a 1, porque conceptualmente, este coeficiente, al igual que los otros coeficientes de confiabilidad, se calcula para responder a la pregunta de qué tan semejante es ese conjunto de datos. Lo que se determina, esencialmente, es la semejanza en una escala que va de 0 (absolutamente no semejante), a 1 (perfectamente idénticos). Debe

tomarse en consideración, que cuando el valor del coeficiente alfa es demasiado alto (mayor a 0.90), ello puede deberse a la existencia de redundancia entre los reactivos, estímulos o preguntas.

Vale la pena poner énfasis en el hecho de que todos los indicadores de confiabilidad, proporcionan un índice que es característico del grupo particular de calificaciones obtenidas en esa prueba, y no de la prueba en sí misma; las medidas de confiabilidad son estimaciones, y éstas están sujetas a error. La cantidad precisa de error inherente en la estimación de la confiabilidad variará de acuerdo con la muestra de respondientes con quienes se haya calculado; la confiabilidad publicada en el manual de la prueba puede ser impresionante, sin embargo, ésta se debe al grupo particular de sujetos con el que se determinó. Si se utiliza a un nuevo grupo de sujetos, muy diferente a aquel con el que se determinó la confiabilidad publicada en el manual, el coeficiente que se obtenga puede no ser tan impresionante, y en ocasiones, hasta puede ser inaceptable.

Lo anteriormente señalado se debe principalmente, a las diferencias culturales que pueden existir entre las respuestas dadas por la muestra de sujetos con la que se desarrolló el instrumento y la de interés actual del investigador, como se señala en la obra de Díaz-Guerrero, al hablar de las premisas histórico socio culturales del mexicano.[10,11]

En términos generales se puede decir que el propósito de establecer el coeficiente de confiabilidad de cualquier instrumento que se use para medir cualquier variable se debe a la naturaleza de las variables por un lado, y las circunstancias que rodean a la aplicación de la prueba, por el otro. El error de medición puede ser de diferentes tipos: errores no identificados, errores en la calificación, errores en la administración, y también errores en la construcción del instrumento.

Comentarios finales

Otras consideraciones que se deben tomar en cuenta se pueden referir a la naturaleza de la prueba: a) los reactivos son homogéneos o heterogéneos: las pruebas diseñadas para medir un solo factor como una habilidad o un rasgo, se espera que sean más homogéneas y por lo tanto con un mayor grado de consistencia interna; b) la característica, habilidad o rasgo medido, es dinámico o estático: si el rasgo fuera dinámico, se esperaría una menor consistencia interna, en respuesta a su modificación o cambio en función de experiencias emocionales o cognoscitivas; c) la prueba es de velocidad (instrumento con reactivos de nivel de dificultad uniforme, que dando suficiente tiempo para responderla, todos los respondientes deberían de completarla de manera correcta), o de poder (cuando a pesar de dar suficiente tiempo para responder a todos los reactivos, siendo algunos tan difíciles que es poco probable que alguno de los sujetos obtenga una calificación perfecta), en el primer caso se obtendría un coeficiente de confiabilidad espuriamente alto; d) el rango de las calificaciones es o no, restringido en virtud del tamaño de la muestra empleada, la confiabilidad será menor en el caso restringido, y mayor en el otro; e) la prueba está relacionada con un criterio o no, y en este caso, la confiabilidad no dice gran cosa, pues lo importante es si el sujeto alcanzó el puntaje criterio prestablecido o no lo alcanzó.

Una pregunta que los investigadores se hacen con frecuencia es la referida a que tan alto o de qué tamaño tiene que ser el coeficiente de confiabilidad. Esto dependerá del propósito e importancia de las decisiones que se tomarán con base en los puntajes obtenidos por los sujetos. La confiabilidad es un atributo necesario en todas las pruebas que se utilicen para evaluar a alguien; en algunas ocasiones se requerirá de un valor más alto y en otras puede no ser tan importante. Si el resultado de las calificaciones obtenidas por los sujetos implica la toma de una decisión de vida o muerte, por supuesto que deberá ser lo más alto que se pueda. Si la prueba se utiliza en combinación con muchas otras, y no forma parte importante del proceso de la decisión que se haya de tomar, puede ser un poco más baja; los coeficientes de 0.90 y más, son los mejores (incluyendo la posibilidad de la existencia de redundancia en los reactivos, señalada anteriormente), equivaldría a un 10 de calificación; los coeficientes de 0.80 y más, serían equivalentes a un 9; de 0.65 a 0.79, sería una calificación de 6 a 7 (“de panzazo”); más bajos de 0.65, son inaceptables.

Cuando una prueba o escala mide una variable compleja (por ejemplo personalidad, inteligencia, salud mental, aprendizaje, etc.) constituida por diversas subvariables, como podrían ser diferentes rasgos de personalidad, o tipos de inteligencia, o diversos síntomas, o conocimientos y habilidades, parte del procedimiento consiste en establecer el número de factores existentes, por medio de alguno de los tipos de análisis factorial existentes. El método más adecuado para establecer la confiabilidad de dicho tipo de pruebas suele ser el Coeficiente Alfa de Cronbach, que se calcula para cada una de las subvariables o factores arrojados por la factorización llevada a cabo. Solo se aceptan como confiables, aquellos factores con valores alfa iguales o mayores a .65, como se señaló anteriormente. Los demás factores, que no hayan alcanzado dicho valor, deberán ser rechazados. Y también se acostumbra a calcular el valor Alfa general, o del total de la prueba.

Referencias

1. Cronbach LJ. Essentials of psychological testing. United States of America, New York: Harper and Row Publishers; 1960.
2. Magnusson D. Teoría de los tests. México: Editorial Trillas; 1969.
3. Nunnally JC, Bernstein IH. Psychometric Theory. United States of America, New York: McGraw-Hill Inc; 1994.
4. Gulliksen H. Theory of mental tests. United States of America, New York: John Wiley and Sons; 1967.
5. Cohen RJ, Swerdlik ME. Psychological testing and assessment. New Delhi: Tata McGraw-Hill Publishing Company Limited; 2005.
6. Anastasi A, Urbina S. Tests psicológicos. México: Prentice Hall; 1998.
7. Anastasi A. Psychological testing. United States of America, New York: The Macmillan Company; 1966.
8. Guilford JP. Psychometric methods. United States of America, New York: McGraw-Hill Book Company; 1954.
9. Nunnally JC. Psychometric theory. United States of America, New York: McGraw-Hill Book Company; 1967.
10. Díaz-Guerrero R. Psicología el Mexicano: Descubrimiento de la etnopsicología. México: Editorial Trillas; 1996.
11. Díaz-Guerrero R. Bajo las garras de la cultura: Psicología del Mexicano 2. México: Editorial Trillas; 2003.