

Metodología de investigación y lectura crítica de estudios

Aproximación al uso del coeficiente alfa de Cronbach

Heidi Celina Oviedo¹, Adalberto Campo-Arias²

Resumen

Introducción: Las escalas se usan frecuentemente en la investigación y en la práctica clínica de la psiquiatría. No obstante, todas las escalas deben ser formalmente validadas antes, con el propósito de conocer de las propiedades psicométricas en una población específica. *Descripción:* Las escalas pueden ser validadas con el uso de un patrón de referencia o sin él; la segunda opción es la única forma de validar algunos instrumentos. Establecer la consistencia interna de una escala es una aproximación a la validación del constructo y consiste en la cuantificación de la correlación que existe entre los ítems que la componen. Valores de alfa de Cronbach entre 0,70 y 0,90 indican una buena consistencia interna. La determinación del alfa de Cronbach se indica para escalas unidimensionales entre tres y veinte ítems. Se debe informar este valor cada vez que se emplee la escala y puede usarse como única medida de validez cuando es difícil una validación con un patrón de referencia.

Palabras clave: confiabilidad, consistencia interna, alfa de Cronbach, escalas.

Title: An Approach to the Use of Cronbach's Alfa

Abstract

Introduction: Scales are frequently used in clinical and research psychiatry. However, scales must be validated in order to know their psychometric properties when they are administered to special populations. *Description:* Establishing the internal consistency of a scale is an approach to validate its construct. That is, the quantification of the correlation among the items includes in a scale. Cronbach' alpha values between 0.70 and 0.90 suggest a good internal consistency. The Cronbach' alpha coefficient may be estimated when a three- to twenty-item one-dimensional scale is administered and particularly when it is difficult to conduct a validation study against a gold standard.

Key words: Reliability, internal consistency, Cronbach's alpha, rating scales.

.....
¹ Médica, asistente de investigación del Grupo de Neuropsiquiatría, Universidad Autónoma de Bucaramanga, Colombia.

² Psiquiatra, profesor asociado de la Línea de Salud Mental, miembro del Grupo de Neuropsiquiatría de Universidad Autónoma de Bucaramanga, Colombia.

Las escalas de medición se usan para cuantificar, por ejemplo, síntomas depresivos, síntomas de ansiedad, síntomas positivos y negativos de la esquizofrenia, autoestima, bienestar, religiosidad o eficacia de una intervención. Sin embargo, no se ha concertado un nombre único para estas medidas; se han llamado *índices, cuestionarios, escalas, inventarios y pruebas* (1). Nosotros las llamaremos en conjunto, simplemente, *escalas*.

Las escalas se encuentran en todas partes, incluyendo internet, ya sea validadas o no, o de autor existente o inexistente, por lo cual, antes de aplicarlas es importante tener pruebas válidas y confiables de su utilidad (2).

No sólo la psiquiatría ha hecho uso de las escalas, también lo han hecho diferentes ramas de la psicología, la educación, la sociología, el mercadeo, la arquitectura y la medicina en general. Incluso, el uso de estos instrumentos es más común y alcanza mayor desarrollo en profesiones no médicas (3). Su empleo se ha popularizado debido a que constituyen, la mayoría de las veces, un ahorro de recursos, su aplicación es práctica y evita riesgos innecesarios, como cuando se realizan pruebas más “invasivas” o cuando no existe un patrón de referencia por la naturaleza del concepto (4).

El principal objetivo de una escala es determinar el valor de una

variable de forma tan precisa como sea posible. Con ello se muestra su utilidad y, por ende, su calidad. Para garantizar la precisión de la medida es necesaria la validación formal (5). Dentro del proceso de validación tenemos dos componentes para que una escala cumpla su objetivo: el primero es la *validez*, que indica si la cuantificación es exacta y, el segundo es la *confiabilidad*, que alude a si el instrumento mide lo que dice medir y si esta medición es estable en el tiempo (6-8). Tanto la validez y la confiabilidad son conceptos interdependientes, pero no son equivalentes. Un instrumento puede ser consistente (tener una gran confiabilidad), pero no ser válido; por eso las dos propiedades deben ser evaluadas simultáneamente siempre que sea posible (6,9).

Debido al gran uso que se hace hoy de las escalas y a la introducción de términos de la *psicometría* (estudio de las medidas de las funciones mentales) en el lenguaje cotidiano, el objetivo del presente artículo es revisar de una manera sencilla el concepto *consistencia interna* y la interpretación del coeficiente alfa de Cronbach.

Confiabilidad

Según la teoría clásica, la confiabilidad se define como el grado en que un instrumento de varios ítems mide consistentemente una

muestra de la población (10). La medición consistente se refiere al grado en que una medida está libre de errores. El coeficiente de confiabilidad se expresa con la letra r e indica la fuerza de la asociación. El valor r varía entre -1 y $+1$, un valor de 0 indica que no existe relación entre los dos puntajes, mientras que un valor cercano a -1 o a $+1$ indica una relación muy cercana, negativa o positiva, respectivamente. Un valor positivo indica que las personas con puntaje alto en una primera aplicación de la escala también puntuarán alto durante la segunda ocasión (11). Una confiabilidad negativa indica un error en el cálculo o una terrible inconsistencia de la escala (12).

El *error* se conceptualiza como la diferencia entre el valor observado y el verdadero o un valor promedio de las mediciones repetidas (13). Sin embargo, por la imposibilidad teórica de calcular directamente la varianza de la puntuación verdadera, la ecuación fue modificada, de tal suerte que la varianza de la puntuación verdadera es igual a 1 menos la varianza del error ($1-S^2$).

Existen tres formas básicas de medir la confiabilidad y todas buscan determinar la proporción de la varianza en una escala. Fundamentalmente, correlacionan los puntajes obtenidos por una escala con los resultados de la reproducción: la confiabilidad prueba re prueba (*test*

retest), la sensibilidad al cambio (considerada, igualmente, como parte de la validez) y la consistencia interna (para la determinación sólo necesita una aplicación del instrumento) (14).

Consistencia interna

Los instrumentos que buscan medir un constructo pueden ser validados en forma indirecta basándose en la relación que muestren los ítems que componen la escala; es decir, presentan una excelente consistencia interna o interrelación entre las preguntas o incisos que hacen parte de la escala. Esto es realizar una validación de un constructo sin un patrón de referencia.

Por ejemplo, no contamos con un estándar de oro para medir la calidad de vida; no obstante, se diseñó una escala que presenta una buena consistencia interna en las distintas poblaciones donde se ha usado, con un valor del alfa de Cronbach adecuado, o sea, los ítems que la componen guardan una buena correlación entre ellos, de tal suerte que se puede concluir, indirecta y casi osadamente, que la escala tiene un constructo válido (15).

Si, por el contrario, se encuentra que los ítems se correlacionan de manera muy pobre, entonces se pueden interpretar los resultados de tres formas: la primera, la escala no

mide el constructo que quiere medir; la segunda, la conceptualización teórica en que está basado el constructo es incorrecta, y la tercera, que el diseño experimental era inadecuado y falló en probar la hipótesis misma (11).

Este método evalúa el error factorial específico y el error por respuesta al azar, los cuales están distribuidos independientemente dentro de los ítems y tienden a cancelarse mutuamente cuando los puntajes de los ítems son sumados (16). La escala es aplicada a los sujetos en un momento puntual y mide, en pocas palabras, la interrelación de los ítems de la escala (15,17), para lo cual la fórmula más usada es el coeficiente alfa de Cronbach (18,19).

Coeficiente alfa de Cronbach

El coeficiente alfa fue descrito en 1951 por Lee J. Cronbach (15). Es un índice usado para medir la confiabilidad del tipo consistencia interna de una escala, es decir, para evaluar la magnitud en que los ítems de un instrumento están correlacionados (3,17). En otras palabras, el alfa de Cronbach es el promedio de las correlaciones entre los ítems que hacen parte de un instrumento (20). También se puede concebir este coeficiente como la medida en la cual algún constructo, concepto o factor medido está presente en cada ítem. Generalmente,

un grupo de ítems que explora un factor común muestra un elevado valor de alfa de Cronbach (3,21).

Hay otras formas de hallar la consistencia interna: la fórmula 20, propuesta en 1937 por Kuder-Richardson (KR-20), es recomendada para medir la consistencia interna de una escala cuando los ítems tienen respuestas dicotómicas; además, equivale, matemáticamente, al coeficiente alfa de Cronbach, de posterior desarrollo (22). El método de Rulon (1939) mide la correlación entre las dos mitades de la escala y generalmente se encuentra en los estudios de validación de la primera mitad del siglo XX (23). El coeficiente de Kristof (1974) se usa para estimar la consistencia interna en una prueba dividida en tres partes muy correlacionadas (24). El coeficiente de Angoff-Feldt (1975) se determina cuando una escala sólo puede ser dividida en dos partes de tamaño arbitrario, pero homogéneas en contenido (25); sin embargo, tiene la limitación de que en algunas ocasiones puede sobreestimar la consistencia interna verdadera (26). El coeficiente beta, propuesto por Raju (1977), pondera la consistencia interna de una escala cuando es dividida en dos o más partes desiguales (27). Cuando falta interrelación en los componentes de una escala, se puede subestimar la consistencia interna (26). El coeficiente de Feldt-Glimer (1983) es una variante para cuando la escala está

dividida en tres o más partes iguales (28). Y así podríamos seguir citando múltiples coeficientes, pero al final la mayoría son variantes del alfa de Cronbach (26). Todas estas formas exploran indirectamente la homogeneidad o la dimensionalidad de una escala, es decir, unidimensional o multidimensional, o cuántos dominios explora la escala (3).

La popularización del coeficiente alfa de Cronbach se debe a la practicidad de su uso, ya que requiere una sola administración de la prueba. Además, tiene la ventaja de corresponder a la media de todos los posibles resultados de la comparación que se hace en el proceso de dividir en mitades una escala (29).

La forma más sencilla de calcular el valor del alfa de Cronbach es multiplicar el promedio de todas las correlaciones observadas en los ítems por el número de ítems que componen una escala, y luego dividir el producto entre el resultado de la suma de 1 más el producto de la multiplicación del promedio de todas las correlaciones observadas por el resultado de la resta de 1 al número de ítems: $\{a = n \cdot p / 1 + p(n-1)\}$, donde n es el número de ítems y p es el promedio de todas las correlaciones (30).

Para una escala de veinte ítems, como la de Zung para depresión o para ansiedad, el número total de

correlaciones entre todos los ítems es 190. Este número se calcula multiplicando el número de ítems por el resultado de dividir entre 2 el número de ítems menos 1, así: $\{n(n-1)/2\}$, donde n es el número de ítems.

Fórmulas más sofisticadas para averiguar la consistencia interna de una escala se calculan a partir de la varianza de cada ítem y la varianza total de la escala: $\{a = n / n-1 (1 - Sst^2 / sT^2)\}$, donde n es el número de ítems, st^2 es la varianza de cada ítem y sT^2 es la varianza total de la escala (31).

Coeficiente alfa estratificado

Una manera de mejorar el valor del coeficiente de consistencia interna cuando las escalas tienen varios dominios es estratificarla en subescalas dependiendo del contenido; de esta forma se puede hallar un mejor valor estimado de la consistencia interna que en el coeficiente alfa no estratificado (32,33).

Coeficiente alfa ponderado

Se usa cuando una escala tiene ítems con diferentes tipos de respuesta. Se le proporciona un peso a cada discrepancia y se supone que los ítems están igualmente ponderados y se realiza una sumatoria de los valores de alfa ponderados para cada ítem (33).

Interpretación del coeficiente alfa de Cronbach

El valor mínimo aceptable para el coeficiente alfa de Cronbach es 0,70; por debajo de ese valor la consistencia interna de la escala utilizada es baja. Por su parte, el valor máximo esperado es 0,90; por encima de este valor se considera que hay redundancia o duplicación. Varios ítems están midiendo exactamente el mismo elemento de un constructo; por lo tanto, los ítems redundantes deben eliminarse. Usualmente, se prefieren valores de alfa entre 0,80 y 0,90 (33). Sin embargo, cuando no se cuenta con un mejor instrumento se pueden aceptar valores inferiores de alfa de Cronbach, teniendo siempre presente esa limitación (3).

Es necesario tener en cuenta que el valor de alfa es afectado directamente por el número de ítems que componen una escala. A medida que se incrementa el número de ítems, se aumenta la varianza sistemáticamente colocada en el numerador, de tal suerte que se obtiene un valor sobreestimado de la consistencia de la escala (12). De igual manera, se debe considerar que el valor del alfa de Cronbach se puede sobreestimar si no se considera el tamaño de la muestra: a mayor número de individuos que completen una escala, mayor es la varianza esperada (19).

Uso del coeficiente alfa de Cronbach

El coeficiente alfa de Cronbach es una propiedad inherente del patrón de respuesta de la población estudiada, no una característica de la escala en sí misma; es decir, el valor de alfa cambia según la población en que se aplique la escala (20). Por ello, en los estudios donde se emplee una escala para cuantificar una característica, por muy popular y validada que sea la escala, debe ser informarse siempre el valor de alfa en esta población y no debemos conformarnos con que sólo se diga que mostró una buena consistencia en otros estudios (34).

El alfa de Cronbach tiene gran utilidad cuando se usa para determinar la consistencia interna de una prueba con un único dominio o dimensión, porque si se usa en escalas con ítems que exploran dos o más dimensiones distintas, aunque hagan parte de un mismo constructo, se corre el riesgo de subestimar la consistencia interna. En estos casos, lo más indicado es calcular un valor de alfa de Cronbach para cada grupo de ítems que componen una dimensión o una subescala (20,32,33) o calcular un coeficiente alfa estratificado (32,33), lo cual significa que este coeficiente se debe emplear para conocer la consistencia interna en escalas unidimensionales, no en escalas multidimensionales (3,20,21,33).

Es importante tener en cuenta que el coeficiente alfa de Cronbach no está indicado para medir la dimensionalidad de una escala; por ello se desaconseja el uso de este coeficiente cuando a simple vista se intuye que una escala está compuesta por más de un dominio o dimensión y un análisis factorial de la escala lo corrobora (33). Como se ve, para conocer la dimensionalidad de una escala se recomienda el uso de un análisis factorial (26).

El coeficiente alfa de Cronbach es más fidedigno cuando se calcula a una escala de veinte ítems o menos. Las escalas mayores que miden un solo constructo pueden dar la falsa impresión de una gran consistencia interna cuando realmente no la poseen (33).

Así mismo, no se puede llegar al extremo de diseñar escalas de un único ítem. Esto hace la medida poco confiable, pues debido al elevado error producto del azar, es improbable que un único ítem abarque la amplitud de un concepto complejo, además tiene la limitación de que sólo puede clasificar a los individuos en dos categorías, es decir, es incapaz de mostrar diferentes grados o matices de un espectro o dimensión (35). Se considera que 3 es el número mínimo ítems para una escala que explore un solo dominio o factor (36).

De la misma forma, el coeficiente alfa de Cronbach no puede apli-

carse a instrumentos que evalúen conocimiento o entrenamiento previo, aunque se esté explorando a manera de preprueba de conocimiento o conducta de entrada que se posee en una única y limitada área de estudio (33).

Conocer la consistencia interna de una escala es un paso anterior (tal vez el primero y el más sencillo) al proceso de validación de un instrumento con un patrón de referencia y para algunos constructos es la única forma posible de validación ante la falta de un estándar de oro válido, seguro, de fácil uso y bajo costo (3,14,33).

Conclusiones

El proceso de validación de una escala es un proceso largo y costoso si se necesita comparar con un patrón de referencia. El coeficiente alfa de Cronbach es la forma más sencilla y conocida de medir la consistencia interna y es la primera aproximación a la validación del constructo de una escala. El coeficiente alfa de Cronbach debe entenderse como una medida de la correlación de los ítems que forman una escala. Está indicada la determinación del alfa de Cronbach en escalas unidimensionales que tengan entre tres y veinte ítems y siempre se debe informar este valor en la población específica donde se empleó la escala.

Bibliografía

1. Streiner DL. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry*. 1993;38:140-8.
2. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *Br Med J*. 2002;324:569-73.
3. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol*. 1993;78:98-104.
4. Cabello JB, Pozo F. Estudios de evaluación de las pruebas diagnósticas en cardiología. *Rev Esp Cardiol*. 1997;50:507-19.
5. Kane MT. Current concepts in validity theory. *J Educ Meas*. 2001;38:319-42.
6. Morgan GA, Gliner JA, Harmon RJ. Measurement Validity. *J Am Acad Child Adolesc Psychiatry*. 2001;40:729-31.
7. Peter JP. Reliability: a review of psychometric basics and recent marketing practices. *J Mark Res*. 1979;16:6-17.
8. Latour J, Abraira V, Cabello JB, López J. Las mediciones clínicas en cardiología: validez y errores de medición. *Rev Esp Cardiol*. 1997;50:117-28.
9. Nelson-Gray RO. DSM-IV: Empirical guidelines from psychometrics. *J Abnorm Psychol*. 1991;100:308-15.
10. Gliner JA, Morgan GA, Harmon RJ. Measurement reliability. *J Am Acad Child Adolesc Psychiatry*. 2001;40:486-8.
11. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281-302.
12. Krus DJ, Helmstadter GC. The problem of negative reliabilities. *Educ Psychol Meas*. 1993;53:643-50.
13. Kupermintz H. On the reliability of categorically scored examinations. *J Educ Meas*. 2004;41:193-204.
14. Blacker D, Endicott J. Psychometric properties: concepts of reliability and validity. In: Rush AJ, Pincus HA, First MB, Zarin DA, Blacker D, Endicott J, et al. *Handbook of psychiatric measures*. Washington: APA; 2002. p. 7-14.
15. Cronbach LJ. Coefficient alpha and the internal structure of test. *Psychometrika*. 1951;16:297-334.
16. Schmidt FL, Le H, Ilies R. Beyond alpha: an empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychol Methods*. 2003;8:206-24.
17. Bland JM, Altman DG. Validating scales and indexes. *Br Med J*. 2002;24:606-7.
18. Cronbach LJ. Coefficient alpha and internal structure of test. *Psychometrika*. 1951;16:297-333.
19. Bland JM, Altman DG. Cronbach's alpha. *Br Med J*. 1997;314:572.
20. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J Pers Assess*. 2003;80:217-22.
21. Rogers WM, Schmitti N, Mullins ME. Correction for unreliability of multifactor measures: comparison of Alpha and parallel forms approaches. *Organ Res Methods*. 2002;5:184-99.
22. Kuder GF, Richardson MW. The theory of the estimation of test reliability. *Psychometrika*. 1937;2:151-60.
23. Rulon PJ. A simplified procedure for determining the reliability of a test of split halves. *Harvard Educ Rev*. 1939;9:99-103.
24. Feldt LS. Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika*. 1975;40:557-561.
25. Raju NS. A generalization of coefficient alpha. *Psychometrika*. 1977;42:549-65.
26. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods*. 2000;5:343-55.
27. Kristof W. Estimation of reliability and true score variance from a split of a test into three arbitrary parts. *Psychometrika*. 1974;39:491-99.
28. Feldt LS, Brennan RL. Reliability. In Linn LR. *Educational measurement*. 3rd ed. New York: Macmillan. 1989; p. 105-46.
29. Kwok WCC, Sharp DJ. A review of construct measurement issues in behavioral accounting research. *J Account Lit*. 1998;17:137-74.
30. Hernández R, Fernández C, Baptista P. *Metodología de la investigación*. 2^{da} ed. México: McGraw-Hill; 1998. p. 376-9.

31. Sánchez R, Gómez C. Conceptos básicos sobre validación de escalas. *Revista Colombiana de Psiquiatría*. 1998;27:21-30.
32. Cronbach LJ, Schoneman P, Mckie D. Alpha coefficient for stratified-parallel test. *Educ Psychol Meas*. 1965;25:291-312.
33. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80:99-103.
34. Knapp TR. Focus on psychometrics: coefficient alpha: conceptualizations and anomalies. *Res Nurs Health*. 1991; 14:457-60.
35. Jaju A, Crask MR. The perfect design: optimization between reliability, validity, redundancy in scale items and response rates. *Am Market Assoc*. 1999;10:127-31.
36. Streiner DL. Figuring out factors: the use and misuse of factor analysis. *Can J Psychiatry*. 1994;39:135-40.

Recibido para publicación: 28 de marzo de 2005

Aceptado para publicación: 15 de julio de 2005

Correspondencia

Adalberto Campo-Arias

Calle 157 No 19-55, Cañaveral Parque

Bucaramanga, Colombia

acampoar@unab.edu.co