

Propiedades Psicométricas de una Escala: la Consistencia Interna

Psychometric properties of a scale: internal consistency

Adalberto Campo-Arias y Heidi C. Oviedo

Instituto de Investigación del Comportamiento Humano, Bogotá, Colombia.
campoarias@comportamientohumano.org

Recibido 9 Abril 2008/Enviado para Modificación 11 Octubre 2008/Aceptado 15 Noviembre 2008

RESUMEN

La confiabilidad tipo consistencia interna se refiere al grado en que los ítemes de una escala se correlacionan entre ellos. La consistencia interna se calcula con la fórmula 20 de Kuder-Richardson para las escalas dicotómicas y el coeficiente de alfa de Cronbach para escalas politómicas. La consistencia interna se considera aceptable cuando se encuentra entre 0,70 y 0,90. Para determinar la consistencia interna de una escala de veinte ítemes se necesitan entre cinco y veinte participantes por cada ítem. La consistencia interna de un instrumento varía según población en que se aplica por ello es siempre necesario informar el valor cada vez que se aplique.

Palabras Clave: Psicometría, escalas, estudios de validación, revisión (*fuentes: DeCS, BIREME*).

ABSTRACT

Internal consistency reliability is the degree of correlation between a scale's items. Internal consistency is calculated by Kuder-Richardson's formula 20 for dichotomous choices and Cronbach's alpha for polytomous items. 0,70 to 0,90 internal consistency is acceptable. 5-25 participants are needed for each item when computing the internal consistency of a twenty-item scale. Internal consistency varies according to population and then it is necessary to report it always that scale is used.

Key Words: Psychometrics, scales, validation studies, review (*source: MeSH, NLM*).

Los instrumentos de medición como las escalas se utilizan para cuantificar atributos, cualidades o propiedades, constructos o conceptos completamente teóricos, que es imposible medir o cuantificar de otra forma (1).

Las escalas, como todos los instrumentos de medición, deben ser plenamente válidos y confiables, es decir, mostrar altos valores de validez y de confiabilidad (2). La validez alude la capacidad del instrumento de medir el constructo que pretende cuantificar (3-5) y la confiabilidad, a la propiedad de mostrar resultados similares, libre de error, en repetidas mediciones (6). Si asume que las condiciones en la cual realiza la aplicación de la escala y del atributo investigado cuantificado se mantienen estables en el tiempo, no ha cambiado (7). La confiabilidad es una condición necesaria; pero, no suficiente para garantizar la validez de un instrumento (8).

Algunas propiedades psicométricas se pueden calcular con la aplicación del instrumento en una sola ocasión en una muestra o población objeto de estudio (9,10). Las características de un instrumento que se pueden explorar con una única aplicación de una escala son la consistencia interna (homogeneidad), la dimensionalidad (estructura factorial o interna) y la confiabilidad del constructo (5,11).

El presente artículo se centrará en los aspectos relevantes de la consistencia interna de los instrumentos de medición que se pueden utilizar en diferentes contextos: ciencias de la salud, ciencias sociales, mercadeo, etc.

Consistencia interna

La confiabilidad tipo consistencia interna se refiere al grado en que los ítemes, puntos o reactivos que hacen parte de una escala se correlacionan entre ellos, la magnitud en que miden el mismo constructo (10,12,13). Si los puntos que componen una escala teóricamente miden el mismo constructo deben mostrar una alta correlación, es decir, la escala debe mostrar un alto grado de homogeneidad (10,14).

La consistencia interna de un instrumento se puede calcular tanto para escala con patrón de respuesta dicotómico como para aquellas con opciones de repuesta politómica. Se proponen diferentes fórmulas para el cálculo de la consistencia interna de los instrumentos de medición (15-19). Estas fórmulas menos usadas en la actualidad guardan el mismo principio de las fórmulas más tradicionales y conocidas en el contexto de la psicometría (20-22). Sin embargo, las más conocidas son la prueba de alfa de Cronbach y la fórmula 20 de Kuder-Richardson (23).

La fórmula 20 de Kuder-Richardson se indica para el cálculo de la consistencia interna de escalas dicotómicas (24) y el coeficiente de alfa de Cronbach para escalas politómicas (25). Estos coeficientes tienen equivalencia matemática (10).

Es importante tener presente que este tipo de coeficientes sólo se puede calcular apropiadamente a escalas que miden atributos o características y no el conocimiento sobre un tópico particular, es decir, no se puede determinar la consistencia interna a una prueba de conocimiento que se aplica en un colegio o universidad, o sea, que necesitan entrenamiento o conocimiento previo en un tópico particular (26).

Aunque, habitualmente se acepta por lo general que la consistencia interna de una escala puede estar en el rango entre 0 y 1 (27). Teóricamente, los valores podrían oscilar entre -1 y 1 (28,29). Valores de 1 indicarían una correlación entre ítemes cercanas a la perfección; valores de 0, ninguna correlación; y coeficiente de -1, una correlación negativa entre los ítemes. Es posible ver coeficientes extrañamente negativos cuando se calcula la consistencia interna para una escala que consta de sólo dos ítemes (27-29).

Cómo calcular la consistencia interna

Existen disponibles un par de fórmulas para calcular la consistencia interna de una escala con patrón de respuesta politómica:

$$\text{Alfa de Cronbach} = k \bar{r} / 1 + \bar{r} (k-1)$$

k , número de ítemes

$$r = \Sigma r / nr$$

r , correlación de Pearson

$$nr = k (k - 1) / 2$$

Una segunda fórmula disponible basa el cálculo del coeficiente en la varianza de los ítemes y la varianza de la puntuación total. Esta fórmula es:

$$\text{Alfa de Cronbach} = k / k - 1 [1 - \Sigma \sigma_i^2 / \sigma_T^2]$$

k , número de ítemes

σ_i , varianza del ítem

σ_T , varianza total de la escala

Los paquetes estadísticos como SPSS o STATA calculan la consistencia interna de una escala dicotómica con el mismo comando que lo hace para una escala politómica, con igual resultado (30, 31), la consistencia interna de una escala dicotómica se calcula:

$$KR-20 = k / k - 1 [1 - \sum p_i q_i / \sigma_T^2]$$

k , número de ítems

p_i , % de afirmativo del ítem

q_i , complemento de p

σ_T , varianza total de la escala

Interpretación de los valores de los coeficientes de consistencia interna

Para los más liberales, la consistencia interna de una escala se considera aceptable cuando se encuentra entre 0,70 y 0,90 (27,32-34). Otros más conservadores sugieren que la consistencia interna de un instrumento es adecuada si el coeficiente alcanza valores entre 0,80 y 0,90, más aún cuando se está en los primeros estadios de construcción de una escala (14,26,35,36).

Algunos conciben la confiabilidad tipo consistencia interna como la proporción de la variabilidad que es debida a la variabilidad real de las puntuaciones observado y no el resultado de algún tipo de error (33). Algún grado de error siempre es posible en cualquier medición (37). Se observa un coeficiente alfa de 0,85 se puede asumir que el 85 de la variabilidad es cierta y el 15 restante probablemente producto del error inherente al tipo de medición, el error no sistemático (10,32).

Valores de consistencia interna inferiores a 0,70 indican una pobre correlación entre los ítems y aquellos por encima de 0,90 indican redundancia o duplicación de ítems, es decir, que por los menos un par de ítems miden exactamente el mismo aspecto de un constructo y uno de ellos debe eliminarse (38).

De la misma forma, coeficientes por encima de 0,90 se pueden observar con regularidad cuando se explora la consistencia interna de instrumentos con más de veinte ítems dado que el coeficiente, si se recuerdan las fórmulas como se calcula, se afecta en forma directa por el número de ítems que hacen parte de la escala, se sobreestima el coeficiente (14,26). Para instrumentos formados por veinte ítems se recomienda que se calcule la consistencia interna de diferentes subescalas, subgrupo de ítems, que generalmente hacen parte de una escala (29).

Debido a la peculiaridad del coeficiente de consistencia interna de afectarse por el número de ítems (39). Cuando se tienen varias escalas que miden el mismo constructo que muestran similares valores de consistencia interna se puede suponer que las correlaciones entre los ítems son mayores en aquella que está compuesta por el menor número de ítems (12). No obstante, lo ideal es comparar escalas que exploren el mismo constructo y tenga el mismo número de ítems (40). El impacto es más marcado en la consistencia interna con la reducción del número de ítems si se ensaya con una muestra relativamente pequeña. Es posible hacer ajuste en el valor de la consistencia interna cuando se comparan dos versiones de una misma escala que difieren sustancialmente en el número de ítems. Siempre y cuando la correlación promedio entre los ítems sea similar. Para ello se puede usar la siguiente fórmula:

$$\alpha^n = m \alpha^a / 1 + (m-1) \alpha^a$$

α^n = nuevo valor de alfa α , si la escala se extiende o acorta.

α^a = valor de α anterior

m= coeficiente de la división del número de ítems de la escala nueva versión entre el número de la anterior.

Es posible observar un incremento en la consistencia con la reducción en el número de ítems cuando se eliminan aquellos que muestran correlaciones corregidas bajas con la puntuación total, es decir, cuando el o los ítems eliminados son de poca calidad y exploran pobremente el constructo (41).

De la misma forma, se debe tener presente cuando se comparan una versión dicotómica con una versión politómica de la misma escala que la confiabilidad se afecta por el número de opciones posibles para un ítem, a mayor número de opciones es mayor el coeficiente de confiabilidad (39).

Tamaño de la muestra en estudio para conocer la consistencia interna

El tamaño de la muestra es muy importante en la mayoría de los estudios cuantitativos. En general, se acepta que las muestras grandes en los estudios reducen en forma significativa la posibilidad de error en el parámetro evaluado (42,43). Sin embargo, para los estudios en que se define la consistencia interna esto no es del todo cierto (42). Si se considera que en la fórmula para el cálculo de la consistencia interna no se considera el número de participación en el estudio sino el número de ítems que hacen parte de la escala o instrumento (24,25). A pesar de esto, los más conservadores, con el propósito de minimizar

los posibles errores debidos al azar, sugieren que este tipo de validación se realice con muestras no menores de 400 participantes (44). Mas, están difundidas recomendaciones más flexibles (41,46). Se acepta para escalas de veinte ítemes que el tamaño de la muestra esté entre cinco y veinte participantes por cada ítem que hace parte de la escala, es decir, entre 100 y 400 sujetos (14,26,46). Cuando el número de participantes calculado para una escala compuesta por menos de diez ítemes es inferior a 100 sujetos, se recomienda que la muestra se forme con diez participantes por cada diez ítemes, como cuando se va a realizar un análisis factorial exploratorio (14,46). Para simplificar este proceso del cálculo del tamaño de la muestra se sugiere que siempre que se pueda se cuente con una muestra de por lo menos cien participantes en este tipo de estudio (14).

Aplicaciones prácticas de los coeficientes de consistencia interna

El proceso de validación de instrumentos es un proceso continuo en constante evaluación que invita a modificaciones que se basen en las investigaciones disponibles (47). Siempre se necesario hacer adaptaciones lingüísticas cuando se traducen escalas de otros idiomas y, aún, en el mismo idioma cuando el contexto cultural es distinto (48). Es importante conocer siempre la consistencia interna de un instrumento en una población específica (22,48). Es frecuente observar estudios en los que se usan escalas y no se informa la consistencia interna de las puntuaciones en la muestra investigada; en cambio, se presenta la consistencia interna que se encontró en otro estudio (24).

Aunque, la consistencia interna es una medida de confiabilidad es un indicador indirecto de la validez en dicha población (48). Se conceptualiza que la consistencia interna no es una propiedad inherente de una escala sino la medida del patrón de respuesta del instrumento en el grupo que responde la escala, es decir, que observar un alto valor consistencia interna para una escala en una población, por ejemplo de adultos jóvenes, no garantiza un coeficiente similar en otra con alguna característica sociodemográfica en común, por ejemplo, en adultos mayores (12,14,49). Igualmente, esto sugiere que los resultados observados pueden indicar la necesidad de adaptaciones de la escala para poblaciones específicas (47,50).

De la misma forma, es necesario tener presente que la consistencia interna es mayor en la medida que existe mayor variación en las puntuaciones en los ítemes individuales y en la escala global (14). Altos grados de variabilidad de

las puntuaciones se logran cuando el instrumento se aplica en una población lo suficientemente heterogénea en relación con el atributo o característica que se intenta medir (6,39).

La consistencia interna es una medida de la correlación existente entre los ítems que componen cualquier instrumento de medición tipo escala que, por general, se puede realizar con una muestra de 100 participantes. Los valores son aceptables cuando son iguales o superiores a 0,70 y menores o iguales a 0,90. Este coeficiente debe calcularse siempre se aplique el instrumento en una población, aunque se haya observado en usos precedentes valores en el rango deseado. La consistencia interna varía según las características de los participantes♣

REFERENCIAS

1. Bland JM, Altman DG. Validating scales and indexes. *Br Med J* 2002; 324: 606-607.
2. Hernández R, Fernández-Collado C, Baptista P. Metodología de la investigación. Cuarta edición. México: McGraw-Hill Interamericana; 2006.
3. Morgan GA, Gliner JA, Harmon RJ. Measurement validity. *J Am Acad Child Adolesc Psychiatry* 2001; 40: 729-731.
4. Rodríguez MA, Lopera J. Conceptos básicos en la validación de escalas en salud mental. *Revista CES Medicina* 2002; 16 (2): 31-39.
5. Lamprea JA, Gómez-Restrepo C. Validez en la evaluación escalas. *Rev Colomb Psiquiatr* 2007; 36: 340-348.
6. Kaplan RM, Saccuzzo DP. Pruebas psicológicas. Sexta edición. México: Thomson; 2006.
7. Rudner L, Schafer W. What teachers need to know about assessment. Washington, DC: National Education Association; 2002.
8. Gliner JA, Morgan GA, Harmon RJ. Measurement reliability. *J Am Acad Child Adolesc Psychiatry* 2001; 40: 486-488.
9. Ferketich S. Internal consistency estimates of reliability. *Res Nurs Health* 1990; 13: 437-440.
10. Hebson RK. Understanding internal consistency reliability estimates: a conceptual primer on coefficient alpha. *Meas Eval Couns Dev* 2001; 34: 177-189.
11. 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*. APA, Washington; 2002 (CD-ROM).
12. Streiner DL. Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *J Pers Assess* 2003; 80: 217-222.
13. Sun W, Chou C-P, Stacy AW, Ma H, Unger J, Gallager P. SAS and SPSS macros to calculate standardized Cronbach's alpha using the upper bound of the phi coefficient for dichotomous items. *Behav Res Methods* 2007; 39: 71-84.
14. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 1993; 78: 98-104.
15. Rulon PJ. A simplified procedure for determining the reliability of a test of split halves. *Harvard Educ Rev* 1939; 9: 99-103.
16. Feldt LS. Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika* 1975; 40:b557-561.

17. Raju NS. A generalization of coefficient alpha. *Psychometrika* 1977; 42: 549-565.
18. Kristof W. Estimation of reliability and true score variance from a split of a test into three arbitrary parts. *Psychometrika*, 1974; 39: 491-499.
19. Feldt LS, Brennan RL. Reliability. In Linn LR. *Educational measurement*. 3rd edition. New York: Macmillan; 1989. p. 105-146.
20. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods* 2000; 5: 343-355.
21. Rogers WM, Schmitt N, Mullins ME. Correction for unreliability of multifactor measures: comparison of Alpha and parallel forms approaches. *Organ Res Methods* 2002; 5: 184-199.
22. Oviedo HC, Campo-Arias A. Aproximación al uso del coeficiente alfa de Cronbach. *Rev Colomb Psiquiatr* 2005; 34: 572-580.
23. Hogan TP, Benjamin A, Brezinski KL. Reliability methods: a note on the frequency of use of various types. *Educ Psychol Meas* 2000; 60: 523-531.
24. Kuder GF, Richardson MW. The theory of the estimation of test reliability. *Psychometrika* 1937; 2: 151-160.
25. Cronbach LJ. Coefficient alpha and internal structure of test. *Psychometrika* 1951; 16: 297-333.
26. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003; 80: 99-103.
27. Bland JM, Altman DG. Cronbach's alpha. *Br Med J* 1997; 314: 572.
28. Krus DJ, Helmstadter GC. The problem of negative reliabilities. *Educ Psychol Meas* 1993; 53: 643-650.
29. Campo-Arias A. Usos del coeficiente de alfa de Cronbach (carta al editor). *Biomedica* 2006; 26: 585-588.
30. SPSS for windows 13.0. Chicago: SPSS Inc.; 2005.
31. STATA 9.0 for windows. College Station: StataCorp LP; 2005.
32. Sánchez R, Gómez C. Conceptos básicos sobre validación de escalas. *Rev Colomb Psiquiatr* 1998; 27: 121-130.
33. Terblanche NS, Boshoff C. Improved scaled development in marketing. An empirical illustration. *Int J Mark Res* 2006; 50: 105-119.
34. Cervantes VH. Interpretaciones del coeficiente de alpha de Cronbach. *Avances en Medicion* 2005; 3: 9-25.
35. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess* 1995; 7: 309-319.
36. Roberts P, Priest H, Traynor M. Reliability and validity in research. *Nurs Stand* 2006; 20: 41-45.
37. 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: 206-224.
38. Jaju A, Crask MR. The perfect design: optimization between reliability, validity, redundancy in scale items and response rates. *Am Market Assoc* 1999; 10: 127-131.
39. Onwuegbuzie AJ, Daniel LG. A framework for reporting and interpreting internal consistency reliability estimates. *Meas Eval Couns Dev* 2002; 35: 89-103.
40. Alsawalmeh YM, Feldt LS. Testing the equality of independent alpha coefficients adjusted for test length. *Educ Psychol Meas* 1999; 59: 373-383.
41. Peterson RA. A meta-analysis of Cronbach's coefficient alpha. *J Cons Res* 1994; 21: 381-391.
42. Kopalle PK, Lehmann DR. Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Org Behav Hum Dec Proc* 1997; 70: 189-197.
43. Duhachek A, Coughlan AT, Iacobucci T. Results on standard error of the coefficient alpha index of reliability. *Market Sci* 2005; 24: 294-301.

44. Gómez C, Sánchez R. Cálculo del tamaño de la muestra en psiquiatría y salud mental. *Rev Colomb Psiquiatr* 1998; 27: 131-142.
45. Charter RA. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *J Clin Exp Neuropsychol* 1999; 21: 559-566.
46. Norman GR, Streiner DL. *Bioestadística*. Madrid: Mosby Doyma Libros; 1996.
47. Sánchez R, Echeverry J. Validación de escalas de medición en salud. *Rev Salud Pública* 2004; 6: 302-318.
48. Rubio-Stipec M, Hicks MHR, Tsuang MT. Cultural factors influencing the selection, use, and interpretation of psychiatric measures. In: Rush AJ, Pincus HA, First MB, Zarin DA, Blacker D, Endicott J, et al. *Handbook of psychiatric measures*. Washington: American Psychiatric Association; 2002 (CD-ROM).
49. Knapp TR. Coefficient alpha: conceptualizations and anomalies. *Res Nurs Health* 1991; 14: 457-460.
50. Streiner DL. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry* 1993; 38: 140-148.