

Comparación Estadística de la Confiabilidad Alfa de Cronbach: Aplicaciones en la Medición Educativa y Psicológica

Statistical Comparison of the Cronbach's Alpha Reliability: Applications in the Educational and Psychological Measurement

César Merino Soto¹

Gary J. Lautenschlager, Ph. D.²

Resumen

El artículo presenta una exposición introductoria sobre el coeficiente alfa (Cronbach, 1951), y los análisis adicionales que se pueden efectuar desde el enfoque de la prueba de hipótesis con este estadístico. Esta prueba de hipótesis de basa en hallar si las diferencias entre los coeficientes alfa obtenidos en muestras dependientes o independientes son estadísticamente significativas. Para ello, uno de los autores (Lautenschlager, 1989) ha creado un programa, ALPHATST, que es muy útil para calcular y obtener la información necesaria que permita al investigador resolver la pregunta "¿son significativamente diferentes los coeficientes alfa hallados en muestras independientes (o dependientes)?". Se exponen algunos ejemplos sobre el uso de este enfoque y así como de sus potenciales aplicaciones en el contexto educativo.

Palabras claves: *evaluación, coeficiente alfa, confiabilidad, consistencia interna, prueba de hipótesis.*

Abstract

This report presents an introductory exhibition on the coefficient alpha (Cronbach, 1951), and the additional analyses that can be made from an hypothesis testing approach with this statistics. This test hypothesis is based on finding out if the differences among the alpha coefficients obtained from dependent or independent samples are statistically significant. For this end, one of the authors (Lautenschlager, 1989) have created a program, ALPHATST, which is very useful to calculate and obtain the necessary information that allows the investigator to solve the question "are the alpha coefficients found in dependent or independent samples significantly different?". We show some examples on the use of this approach as well as of their potential applications in the educational context.

Keyword: *assessment, alpha coefficient, reliability, internal consistency, testing hypothesis.*

¹ Psicólogo, Defensoría Municipal del Niño y del Adolescente, Perú. email: sikayax@yahoo.com.ar

² Profesor de Psicología, Department of Psychology, University of Georgia. USA. email: garylaut@arches.uga.edu

Introducción

En el contexto Latinoamericano no se ha difundido apropiadamente la mirada de métodos para analizar aspectos claves de las pruebas que, por ejemplo, se aplican en el área educativa; de esta manera, en lo que concierne al tema de la confiabilidad, se tienen estimaciones alternativas de la consistencia interna derivadas del análisis factorial (Carmines y Zeller, 1979; Heise y Bohrnstedt, 1970), estimaciones que hacen frente a bajos coeficientes alfa (Roberts, Onwuegbuzie y Eby, 2001), estimaciones para una batería de pruebas (Raju, 1977), etc..

Aún con un emergente impulso en la difusión de tales métodos, la carencia de programas computarizados específicos que den agilidad a los cálculos es una limitación adicional. Una de estas situaciones es la del examen de variación de la consistencia interna, mediante la comparación de los coeficientes obtenidos de diferentes muestras, en diferentes momentos o por diferentes métodos de evaluación. El presente documento pretende, en primer lugar dar una visión sintética de la metodología para comparar estadísticamente las estimaciones de confiabilidad; luego, se describirán ejemplos de uso en el análisis psicométrico en el contexto educativo y, finalmente, se describirán los pasos para el uso de un programa diseñado especialmente para este fin.

Desarrollo de la Prueba de Hipótesis para el coeficiente alfa de Cronbach

Un investigador inquieto podría preguntar, luego de haber obtenido las estimaciones de confiabilidad de una prueba de matemáticas en estudiante varones y mujeres: “¿Los coeficientes de confiabilidad que he calculado, son significativamente diferentes?”. Cuando se obtienen las estimaciones de confiabilidad por medio del coeficiente alfa, pueden surgir preguntas que son abor-

dados desde procedimientos inferenciales conocidos (Muñiz, 1998). Los aspectos del cálculo, así como las derivaciones matemáticas, son expuestos en detalle en Muñiz (1998) y Feldt, Woodruff y Salih (1987). En los siguientes párrafos, presentaremos las formulaciones utilizadas por el programa ALPHATST, pero no se seguirá el desarrollo matemático de los mismos y dejamos al lector interesado su iniciativa para explorar en las fuentes bibliográficas originales.

En principio, los procedimientos sobre los que se basa ALPHATST surgen del desarrollo de la teoría de muestreo del coeficiente alfa bajo la presunción de normalidad de los componentes del puntaje de error y verdadero, que fueron creados independientemente por Kristof (1963) y Feldt (1965):

$$F = \frac{1 - \alpha}{1 - \alpha'} \quad (1)$$

Esta prueba estadística se dirige a contrastar el coeficiente alfa obtenido en la muestra frente a un valor poblacional respectivo o hipotetizado. Se debe tener en mente que se asume que el coeficiente obtenido de una muestra específica es uno que representará a la población del cual proviene tal muestra.

Posteriormente, Feldt (1969) elaboró una prueba para probar la igualdad de los coeficientes alfa de dos muestras independientes, expuesta como:

$$w = \frac{1 - \alpha_1}{1 - \alpha_2} \quad (2)$$

\square_1 = coeficiente alfa de la muestra 1

\square_2 = coeficiente alfa de la muestra 2

Hakstian y Whalen (1976) derivaron una extensión hacia la comparación de más de dos coeficientes alfa extraídos de muestras independientes, que produce resultados similares a la derivación estadística creada por Woodruff y Feldt (1986):

$$UX_1 = \frac{\sum_{i=1}^k \left[(1 - \alpha_i)^{-1/3} - \bar{u} \right]^2}{\bar{S}^2} \quad (3)$$

UX_1 = estadístico distribuido aproximadamente como χ^2 , con K-1 grados de libertad; K = número de las muestras independientes, α_i = coeficiente alfa para cada muestra i.

$$\bar{u} = \text{esta definido por } \sum_{i=1}^k \frac{(1 - \alpha_i)^{-1/3}}{K}$$

$$\bar{S}^2 = \text{es definido como } \bar{S}^2 = \sum_{i=1}^k \frac{S_i^2}{K}$$

Para la formulación última,

$$\bar{S}^2 = \frac{2}{9(\tilde{N}_1 - 1)(1 - \alpha_i)^{2/3}}$$

y

$$\tilde{N}_i = \frac{N_i(n_i - 1)}{n_i + 1}$$

Con N_i = número de sujetos en cada muestra i, n_i = número de ítemes en cada prueba.

Para la situación, nada extraña en la investigación educacional, en que se tienen dos coeficientes alfa de muestras dependientes, Feldt (1980) propuso el siguiente estadístico:

$$t = \frac{(\alpha_1 - \alpha_2) \sqrt{N - 2}}{\sqrt{4(1 - \alpha_1)(1 - \alpha_2)(1 - \rho_{12}^2)}} \quad (4)$$

en que:

t = es el valor distribuido con N-2 grados de libertad, N = número de personas en las muestra, α_1 y α_2 = coeficientes alfa de las pruebas, ρ_{12} = correlación entre los sujetos en las dos pruebas

Finalmente, Woodruff y Feldt (1986) extendieron esta técnica a más de dos coeficientes alfa relacionados, expresado como el siguiente estadístico de contraste:

$$UX_2 = \frac{\sum_{i=1}^k \left[(1 - \alpha_i)^{-1/3} - \bar{u} \right]^2}{\bar{S}^2 - \bar{S}_{jk}} \quad (5)$$

en donde:

K = Número de pruebas, N = número de sujetos de la muestra, UX_2 = estadístico distribuido aproximadamente como χ^2 con K- 1 grados de libertad. α = valor del coeficiente alfa de la muestra. Y

$$\bar{S}^2 = \sum_{i=1}^k \frac{S_i^2}{k}$$

, cuyos componentes son:

$$S_i^2 = \frac{2}{9(\tilde{N} - 1)(1 - \alpha_i)^{2/3}}$$

$$\tilde{N} = \frac{N(\tilde{n} - 1)}{\tilde{n} + 1}$$

$$\tilde{n} = \frac{K}{\sum_{i=1}^k 1/n_i} \quad (\text{media armónica del tamaño de las pruebas})$$

$$S_{jk} = \frac{2\rho_{jk}}{9(\tilde{N} - 1)(1 - \alpha_j)^{1/3}(1 - \alpha_k)^{1/3}}$$

$$\text{y } \overline{S_{jk}} = \frac{\sum S_{jk}}{K(K - 1)/2}.$$

Obsérvese que la formulación de esta prueba estadística exige los valores de ρ_{jk} correlación entre las pruebas.

De estos desarrollos metodológicos, podemos establecer que el enfoque de la estadística inferencial tomó su lugar en la metodología para el examen de la consistencia interna.

Otros desarrollos recientes de pruebas de hipótesis para el coeficiente alfa han sido desarrollados por Alsawalmeh y Feldt (1994, 1999), para las situaciones en que el grado de velocidad para terminar la prueba y los límites de tiempo de cada prueba no son comparables entre sí.

Muñiz (1998) expuso que las situaciones en los que se aplicarían los procedimientos inferenciales considerando el coeficiente alfa se dirigen a: 1) establecer el intervalo de confianza entre el que se encontraría el valor poblacional del coeficiente alfa (recuérdese

que la estimación de la confiabilidad realmente se efectúa sobre una muestra) y probar la significancia estadística del coeficiente alfa proveniente de la muestra; 2) comparar coeficientes alfa de muestras independientes, y c) comparar coeficientes alfa de muestras dependientes.

Para cualquiera de los procedimientos mencionados, se considera que la extracción aleatoria de ítems y personas, desde una población de ítems y de sujetos respectivamente, es una presunción que se cumple (Knapp, 1991) al menos teóricamente.

Usos potenciales.

Uno de los usos, quizás el más representativo, de la comparación de las estimaciones de consistencia interna es la investigación del sesgo a nivel de la prueba, específicamente en el estudio del sesgo de la validez de constructo (Reynolds, 2000).

En el contexto de la investigación del sesgo, la comparación de coeficientes alfa ha tomado un lugar en el estudio del sesgo en la validez de constructo (Reynolds, 2000), en el que el supuesto básico que se investiga es la equivalencia de la exactitud de la medición en los grupos de comparación (Reynolds, 2000). Se espera que la varianza de error debe ser constante en los grupos estudiados y que, en consecuencia, si existe una discrepancia estadísticamente significativa en las estimaciones de confiabilidad de ambos grupos, entonces ello debería indicar que la prueba evidencia sesgo en la estimación del puntaje verdadero, aún teniendo los grupos puntajes obtenidos similares (Reynolds, 2000).

En la investigación intercultural (cross-cultural), una situación típica es el uso de un instrumento que proviene de un contexto con diferencias idiomáticas. Si un investigador educacional hispano quiere adaptar un instrumento existente desarrollado

para hablantes ingleses (por ejemplo, una batería de autoconcepto académico para estudiantes del nivel secundario), puede ser una adecuada estrategia comparar el coeficiente alfa de la muestra inglesa con aquel obtenido de la traducción española aplicada a la muestra hispana, siendo esta muestra comparable en edad, género y otros antecedentes. Para esta estrategia, no es una redundancia decir que la muestra hispana debería emparejarse tan cercanamente como sea posible en aquellos factores identificados como importantes para el constructo que se está midiendo.

Para agregar más a lo anterior, la prueba de hipótesis sobre el coeficiente alfa está enfocado más específicamente sobre los aspectos de contenido de la validez. Esto es consistente con el desarrollo del coeficiente alfa, el cual está basado sobre el modelo del muestreo de dominio de Cronbach (Cronbach, 1951). Como se sabe, la validez de contenido es, per se, también parte de la evidencia de la validez de constructo.

A nivel del ítem, el estudio de la influencia de los componentes de las pruebas sobre la confiabilidad puede ser examinada formalmente por una prueba de hipótesis. Por ejemplo, los efectos del número óptimo de alternativas en los ítems ha sido investigado en pruebas de rendimiento de opción múltiple (Trevisan, Sax y Michel: 1991; 1994) y en escalas de actitudes (Trevisan y Paulson, 1996), mediante la prueba de hipótesis para alfa en muestras independientes. Por otro lado, el uso de los intervalos de confianza del coeficiente alfa de Cronbach tuvo un lugar en Barnette (2000), quien teniendo como criterio a la consistencia interna, examinó las variaciones producidas por la inclusión de ítems con el contenido orientado negativamente y las opciones orientadas en sentido inverso. Sus hallazgos fueron significativamente estadísticos y revelaron las precauciones de utilizar ítems negativamente formulados (Barnette, 2000).

Un reciente uso de la prueba de hipótesis para el coeficientes alfa se observa, por ejemplo, en Merino (2003), que comparó los coeficientes alfa provenientes de la aplicación de una prueba de habilidades cognitivas a preescolares, tomando en cuenta el sexo (varones versus mujeres) y ubicación del centro educativo preescolar (urbano versus periurbano). En tal investigación, los resultados apuntan a una relativa influencia de los factores asociados a la ubicación de centros educativos preescolares sobre el decremento del coeficiente alfa en la mitad de las subpruebas; es decir, los preescolares en los colegios periurbanos tendieron a producir coeficientes alfa inferiores a los preescolares aprendiendo en colegios de zona urbana.

Como análisis preliminar de estudios con objetivos no psicométricos, la prueba de hipótesis del coeficiente alfa puede verificar si las diferencias entre ellas puede ser un factor espúreo que debe ser considerado en la interpretación de los resultados. Por ejemplo, Hill (2001) analizó preliminarmente la confiabilidad entre sujetos de etnias distintas (afro-americanos y euro-americanos) antes del análisis por regresión entre etnicidad, ingresos económicos y prácticas de crianza, y el aprestamiento escolar.

Aún con los avances y críticas que ha tenido la estimación de la confiabilidad por medio del coeficiente alfa, la metodología para su análisis parece extenderse favorablemente. De este modo, un enfoque de prueba de hipótesis (para una muestra, para muestras dependientes o independientes) es particularmente útil cuando las evidencias estadísticas sirven para respaldar decisiones. Sin embargo, su utilidad queda tamizada por el entendimiento que el usuario tiene de la estadística inferencial y de la teoría clásica de los test. El lector interesado debería revisar a Muñiz (1998) y Nunnally y Bernstein (1995).

Aplicaciones informáticas.

Lautenschlager (1989) desarrolló un programa (ALPHATST) que hace los cálculos necesarios para los procedimientos inferenciales sobre el coeficiente alfa descritos en las secciones anteriores. Para el ingreso de los datos que se analizarán, se requiere unos comandos escritos en el formato del lenguaje fortran.

Se describen a continuación 3 situaciones que usualmente se presentan en la medición educacional y que el investigador debe resolver. Es de observar que estas situaciones no agotan otras circunstancias en la que se aplica la prueba de hipótesis, pero servirá para ilustrar el uso de los comandos y la interpretación de los resultados. Los detalles de cómo operar con los comandos fortran se enviarán a solicitud del lector interesado.

Las situaciones son las siguientes:

Situación 1

Muestras independientes utilizando la misma evaluación.

Situación 2

Muestra dependientes utilizando diferentes métodos de evaluación.

Situación 3

Muestras dependientes utilizando el mismo método pero con longitudes diferentes.

Ejemplos:

Situación 1. Muestras independientes utilizando la misma evaluación

Una investigadora quiere determinar si una prueba diseñada para medir la autoestima de los niños y que fue elaborada utilizando un grupo de niños de mayor edad, podría también ser utilizada con niños de menor edad. El instrumento tiene un total de 20 ítems. De esta manera, la investiga-

dora administra la prueba de autoconcepto a grupos de niños de 3er grado y 6to grado de primaria, y primer año de secundaria.

Hay 64 niños en la muestra de 3er grado y el valor del coeficiente alfa es 0,72. En 6to grado hay 87 niños y el valor alfa alcanzado es 0,81. En primero de secundaria hay 101 alumnos y su valor alfa es 0,87. ¿Hay alguna diferencia en las confiabilidades para estos grupos de niños?

Los resultados en el archivo concept.out deberían aparecer como:

```
test for alpha coefficients
input file for this run was: concept.txt
# of independent groups: 3
chi-squared value is : 10.8908
df = 2 p < 0,00469
```

Los resultados indican a la investigadora que debería rechazar la hipótesis de que el instrumento de autoconcepto es igualmente confiable a través de los grados estudiados. Parece que la escala es menos confiable para los niños de menor edad. La investigadora puede querer considerar aspectos tales como el nivel de lectura o el formato de respuesta, que puede hacer que la escala sea menos apropiada para los niños más pequeños.

La investigadora, además, podría estar interesada en la comparación entre dos grupos específicos, para ver si difieren en la confiabilidad. Entonces, podría usar solo dos de los grupos para el análisis. Si se realiza el análisis anterior, los resultados aparecerán como un chi cuadrado de 3,0175, g.l. = 1, $p < 0,07847$. Ya que $p < 0,07847$ no es menor a 0,05, la investigadora puede asumir que la prueba es casi igualmente confiable para los grupos examinados, en este caso, los de mayor edad.

Situación 2. Muestras dependientes utilizando diferentes métodos de evaluación.

En este ejemplo, un profesor trabajando en una oficina del Ministerio de Educación quiere determinar si un cuestionario hecho de preguntas breves, que requiere de respuestas del tipo Sí o No, se puede utilizar en lugar de una entrevista estandarizada para obtener información sobre aspectos importantes en la selección de docentes. El cuestionario tiene 60 preguntas y toma 15 minutos completarlo. La entrevista produce calificaciones sobre 10 escalas relacionadas a características personales y también consume 15 minutos su administración. La entrevista cuesta más porque consume un importante tiempo del equipo de personas contratadas para las entrevistas a profesores. ¿Se puede utilizar el cuestionario en vez de la entrevista?

Podría ser útil comparar las confiabilidades de los dos formatos de evaluación (cuestionario y entrevista). Cada método se aplica a la misma muestra de individuos. Aquí, las estimaciones de confiabilidad son estadísticamente independientes. También es importante considerar que las mediciones toman el mismo monto de tiempo para responderlas y que por lo tanto son comparables como evaluaciones equivalentes de la conducta de interés (ver Feldt, Woodruff y Salih, 1987, p. 98, para detalles adicionales).

Supongamos que el cuestionario tiene una consistencia interna de 0,85, las calificaciones de la entrevista 0,81 y la correlación entre ellas es 0,80, para una muestra de 100 sujetos que resolvieron ambas formas de evaluación. (Tome nota que el investigador debería también preocuparse sobre la validez mediante algún procedimiento, aunque ello es un tema aparte. En este ejemplo, una correlación sustancial entre los dos enfoques para medir la ansiedad estado podría proporcionar una evidencia de validez).

Los resultados generados por alphasst exponen un chi cuadrado de 3,3865, g.l. =1, $p < 0,06220$. Asumiendo un valor p de 0,05, el valor chi cuadrado no es significativo para lograr el rechazo de la hipótesis nula. Esto sugiere que las estimaciones de confiabilidad para las dos medidas son aproximadamente iguales. Además, la relativa alta correlación entre ellas podría sugerir que el cuestionario potencialmente puede reemplazar a la entrevista.

Situación 3. Muestras dependientes utilizando el mismo método pero con longitudes diferentes.

Esta descripción es muy similar a la Situación #2, pero implica comparar los mismos tipos de evaluación para abreviar su extensión. Una investigadora educacional quiere determinar la influencia de la reducción del tamaño de una prueba estandarizada. La prueba actual tiene 100 ítems, pero el investigador quiere acortarlo hasta 25 ítems. Ella no quiere perder mucha confiabilidad así que puede decidir utilizar una prueba algo menos extensa.

Ella administra la prueba completa de 100 ítems a una muestra de 200 estudiantes. Entonces, elige subgrupos específicos de cantidad decreciente, como 75, 50 y 25 ítems en cada agrupación y calcula el coeficiente alfa basado en cada grupo tales ítems. Luego, correlaciona los puntajes entre cada una de las versiones.

Asumamos que las confiabilidades de las cuatro longitudes de la prueba, desde el más extenso (100 ítems) al más breve (25 ítems) son: 0,925, 0,907, 0,883 y 0,850 respectivamente. El investigador correlaciona los puntajes de estas pruebas y encuentra la siguiente matriz de correlación:

Long. prueba (# de ítems)	100	75	50	25	
	100	1.00			
	75	0,85	1.00		
	50	0,75	0,70	1.00	
	25	0,65	0,62	0,58	1.00

Los resultados que aparecerán serán:

```

test for alpha coefficients
input file for this run was: case3.txt
# of dependent alphas: 4
chi-squared value is : 24.6245
df = 3 p < 0,00009
    
```

Debido que el valor p es menos que 0,05, el investigador podría concluir que las diferentes formas no tienen igual confiabilidad. Aún permanece la pregunta sobre cuáles de las formas poseen confiabilidades desiguales. Para responder esto uno puede querer remover secuencialmente las versiones abreviadas, una a la vez, y determinar si aún aparecen diferencias. O, alternativamente, uno puede hacer numerosas comparaciones de por pares en análisis separados, que incluyan dos versiones específicas.

Comentarios finales sobre la comparación de las estimaciones de confiabilidad.

En cada una de los casos analizados, se han aplicado ciertos presupuestos estadísticos, los cuales han sido detalladamente resumidos en Feldt, Woodruff y Salih, (1987). Considerando el tamaño de la muestra, un

mínimo de 20 casos por grupo se requiere para la Situación 1, en la que se prueba la hipótesis para muestras independientes. Para las Situaciones 2 y 3, en donde se tiene la misma muestra o muestras emparejadas, se debería trabajar con un mínimo de 50 casos (o pares). Estos requerimientos mínimos del tamaño de la muestra se basan en el trabajo reportado por Feldt y sus colegas.

En situaciones en que las decisiones tienen resultados críticos, la exigencia de una elevada confiabilidad es una preocupación seria. Sin embargo, en el contexto de la investigación y de particulares aplicaciones, no siempre se requerirá un instrumento altamente confiable, de tal modo que la evidencia estadística de las diferencias en la confiabilidad no debería por sí mismo reemplazar el buen juicio del evaluador.

Disponibilidad

El programa está disponible sin costo para los interesados. Solicitarlo electrónicamente a garylaut@arches.uga.edu o sikayax@yahoo.com.ar. El módulo ejecutable, junto con instrucciones completas para el uso de los comandos FORTRAN, será enviado por correo electrónico como archivo adjunto. Alternativamente, solicitar el programa o copias del artículo a César Merino Soto, Enrique Palacios 430, Chorrillos, Lima 9, Perú.

Referencias Bibliográficas

- ALSAWALMEH, Y. M. Y FELDT, L. S. (1994) "A modification of Feldt's test of the equality of two dependent alpha coefficients". *Psychometrika*, 59, Nro. 1, 49-57.
- ALSAWALMEH, Y. M. Y FELDT, L. S. (1999) "Testing the equality of independent alpha coefficients adjusted for test length", *Educational and Psychological Measurement*, 59, Nro. 3, 373-383.

- BARNETTE, J. J. (2000) "*Effects of stem and likert response option reversals on survey internal consistency: If you feel the need, there is better alternative to using those negatively worded stems*". Educational and Psychological Measurement, 60, 361-370.
- BRIERE, J. & RUNTZ, M. R. (1988) "Multivariate correlates of childhood psychological and physical maltreatment among university women", Child Abuse & Neglect, 12, 331-341.
- CARMINES, N. R. y ZELLER, R. A. (1979) Reliability and validity assessment. Londres: Sage.
- CRONBACH, L. J., (1951) "Coefficient alpha and the internal structure of test". Psychometrika, 16, 3, 297-334.
- FELDT, L. S. (1965) "The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty", Psychometrika, 30, 357-370
- FELDT, L. S. (1969) "A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests", Psychometrika, 34, 363-373.
- FELDT, L. S. (1980) "A test of the hypothesis that Cronbach's alpha coefficient is the same for two tests administered to the same sample", Psychometrika, 49, 99-105.
- FELDT, L. S., WOODRUFF, D. J. & SALIH, F. A. (1987), "Statistical inference for coefficient alpha", Applied Psychological Measurement, 11, 1, 93-103.
- HAKSTIAN, A. R. & WHALEN, T. E. (1976) "A K-sample significance test for independent alpha coefficients", Psychometrika, 41, 2, 219-231.
- HEISE, D. R. & BOHRNSTEDT, G. W. (1970) Validity, invalidity, and reliability. In E.E. Borgata y G. W. Bohrnstedt (Eds.), Sociological methodology (pp. 104-129). San Francisco: Jossey-Bass.
- HILL, N. E. (2001) "Parenting and academic socialization as they relate to school readiness: The roles of ethnicity and family income, Journal of Educational Psychology, 93, 4, 686-697.
- KNAPP, T. R. (1991), "Coefficient Alpha: Conceptualizations and Anomalies", Research in Nursing & Health, 1, 457-460
- KRISTOF, W. (1963) "*The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts*", Psychometrika, 28, 221-238.
- LAUTENSCHLAGER, G. J. (1989) "ALPHATST: Testing for differences in values of coefficient alpha", Applied Psychological Measurement, 13, 284.
- MERINO, C. (2003) Análisis de confiabilidad de una prueba cognitiva para preescolares. Documento no publicado.
- MUÑIZ, J. (1998) Teoría clásica de los test (5ta ed.), Madrid: Pirámide.
- NUNNALLY, J. C. y Bernstein, I. J. (1995). Teoría psicométrica (3ra ed.). México, D. F: McGraw-Hill.
- RAJU, N. S. (1977) "A generalization of coefficient alpha", Psychometrika, 42, 549-565.
- REYNOLDS, C. R. (2000) Methods for detecting and evaluating cultural bias in neuropsychological tests, en L. S. Fletcher-Janzen, L. S. Strickland y C. R. Reynolds (Eds.), Handbook of Cross-Cultural Neuropsychology (pp. 249-285). New York: Kluwer Academic / Plenum Publishers.
- ROBERTS, J. K., Onwuegbuzie, A. J. y Eby, J. R. (2001) "The introduction of a measure of instrument homogeneity for interpreting low reliability". Ponencia presentada en la reunión anual de la American Educational Research Association, Seattle, WA, Abril 14.

- TREVISAN, M. S. & PAULSON, F. L. (1996) "The Tversky condition applied to rating scales". *Psychological Reports*, 78, 891-898.
- TREVISAN, M. S., SAX, G. & MICHAEL, W. B. (1991) "The effects of the number of options per item and student ability on test validity and reliability". *Educational and Psychological Measurement*, 51, 829-836.
- TREVISAN, M. S., SAX, G., & MICHAEL, W. B. (1994) "Estimating the optimum number of options per item using an incremental option paradigm". *Educational and Psychological Measurement*, 54, 1, 86-91.
- WOODRUFF, D. J., & FELDT, L. S. (1986) "Test for equality of several alpha coefficients when their sample estimates are dependent", *Psychometrika*, 51, 3, 393-413.

Fecha Recepción Artículo: 19 de Agosto 2003
Fecha Evaluación Final: 30 de Octubre 2003