Avances en Medición, 6, 131-141
2008

COEFICIENTE ALFA Y DIMENSIONALIDAD EN LOS TESTS PSICOMÉTRICOS[1]

José Ant. López Pina[2]
Universidad de Murcia

Resumen

Este trabajo de simulación Monte Carlo estudia el efecto de la multidimensionalidad del test en el coeficiente alfa. Para ello, se han generado matrices unidimensionales bajo la familia de modelos de Rasch (dicotómico y politómico) que han sido la línea base de comparación de otras matrices multidimensionales generadas en distintas condiciones experimentales: formato del ítem (dicotómico y politómico), intervalo de dificultad de los ítems del test, porcentaje de ítems en la segunda dimensión (relevancia de la dimensión), correlación entre las dimensiones (desde no correlación hasta correlación elevada) y el grado de centralidad de la distribución de habilidad. Los resultados muestran que cuando el test es bidimensional, el coeficiente alfa es siempre más bajo que el coeficiente alfa de la línea base (unidimensional). Además, conforme se incrementó el porcentaje de ítems de la segunda dimensión, el coeficiente alfa decreció en todas las condiciones experimentales, pero el coeficiente alfa se acercó al de la línea base (unidimensional) conforme aumentó la correlación entre las dos dimensiones.

Palabras clave: Simulación Monte Carlo, multidimensionalidad, modelo de Rasch, coeficiente alfa.

Abstract

This paper studies the effect of the multidimensionality of test on the alpha coefficient. In order to do this, unidimensional matrices were generated, with the family of Rasch models as a baseline for comparing the multidimensional matrices in various experimental conditions. Format of the items, difficulty interval of the test, percentage of items in the second dimension (relevance of dimension), correlation between dimensions (from no correlation to high correlation) and the degree of centrality in ability distribution were manipulated. The results showed that when the test is two-dimensional the alpha coefficient is always lower than the alpha coefficient of the (unidimensional) baseline. Furthermore, as the percentage of items in the second dimension increased, the alpha coefficient decreased in all experimental conditions, although as the correlation between the two dimensions increased, the alpha coefficient tended towards the (unidimensional) baseline.

Keywords: Monte Carlo simulation, multidimensionality, Rasch model, coefficient alpha.

Introducción

La fiabilidad de las puntuaciones en un tests psicométrico se puede evaluar a través de tres procedimientos: test-retest, formas paralelas y dos mitades (Crocker y Algina, 1986; Gulliksen, 1950; Lord y Novick, 1968). Aunque algunos investigadores emplean el procedimiento test-retest para obtener evidencias de la estabilidad de las puntuaciones, y con mucha menos frecuencia, se construyen formas paralelas para estudiar su equivalencia, el procedimiento de las dos mitades, y su límite, el coeficiente alfa (Cronbach, 1951) se ha convertido en un estándar en cualquier investigación experimental como medio de evaluar la fiabilidad de las puntuaciones.

El coeficiente alfa (Cronbach, 1951) es una expresión del promedio de las covarianzas entre los ítems de un test, cuando las varianzas de error de los ítems son iguales, es decir, cuando los ítems son

[1] Este estudio ha sido financiado por el proyecto 03109/PHCS/05 de la Fundación Séneca (Murcia).

[2] Dirección de contacto: Facultad de Psicología. Campus de Espinardo. Apdo. 4021. 30100-Murcia. Tlf.: 968 363478. Fax: 968 364111. E-mail: jlpina@um.es.

estrictamente paralelos. Si las varianzas de error de los ítems difieren, entonces el coeficiente alfa es más bajo que el promedio de todos los coeficientes de fiabilidad para las dos mitades de un test estimados a través de la ecuación de Spearman-Brown (Cortina, 1993; Miller, 1995). La expresión del coeficiente alfa es:

$$\alpha \leq \frac{n}{n-1}\left(1-\frac{\sum \sigma_j^2}{\sigma_X^2}\right) \qquad (1)$$

donde *n* es el número de ítems, σ_j^2 es la varianza del ítem *j*, y σ_X^2 es la varianza total. Cuando los ítems son dicotómicos, el coeficiente alfa equivale a KR_{20} (Kuder y Richardson, 1937), cuya expresión matemática es:

$$KR_{20} = \frac{n}{n-1}\left(1-\frac{\sum p_j q_j}{\sigma_X^2}\right) \qquad (2)$$

donde p_j es la proporción de personas que acierta o contesta positivamente el ítem y q_j es la proporción de personas que falla o contesta negativamente el ítem.

Unidimensionalidad de las respuestas en un test psicológico

El tópico de la undimensionalidad en los tests ha sido una constante en la investigación psicométrica, ya que la regla de escalamiento que vincula las respuestas de las personas con la puntuación observada, obtenida como la suma no ponderada de las respuestas acertadas en ítems dicotómicos o las categorías marcadas en ítems politómicos, en el rasgo psicológico es desconocida (Suen, 1990). La complejidad de la estructura y formulación de los ítems en un test psicométrico parecen inducir la idea de que difícilmente pueden ser unidimensionales. Así, la puntuación observada que surge de las respuestas a los ítems puede llegar a ser una mezcla de una dimensión esencial (Stout, 1987) o suficiente (Hambleton y Swaminathan, 1985) que explica la variación de las respuestas, y otras dimensiones necesarias para obtener esas respuestas, o de varias dimensiones suficientemente importantes para que se tengan en cuenta a la hora de evaluar qué atributos determinan las respuestas de las personas.

En la investigación psicométrica, la estrategia habitual ha sido emplear el Análisis Factorial Exploratorio (AFE) y en menor cuantía el Análisis Factorial Confirmatorio (AFC) como técnicas capaces de desentrañar o confirmar la dimensionalidad desconocida en los tests, a través de estudiar los patrones de covarianzas/correlaciones entre los ítems que los componen. Dado que el coeficiente alfa es una expresión del promedio de las covarianzas entre los ítems cuando son estrictamente paralelos, también ha sido considerado como una medida de la saturación del primer factor (Hattie, 1985), siempre y cuando se calcule a partir de una matriz de correlaciones. En este caso, Kaiser (1968) mostró que si las correlaciones entre los ítems son iguales a la correlación promedio de los ítems, el coeficiente alfa se relaciona directamente con el primer componente no rotado de una solución de componentes principales.

En la práctica, sin embargo, las puntuaciones provenientes de las respuestas de las personas a los ítems de un test pueden depender de una o más dimensiones, por lo que las covarianzas entre los ítems pueden ser altas aún proviniendo de más de una dimensión, siempre que estas dimensiones intervengan en el proceso de respuesta. Así, el coeficiente alfa será, generalmente, elevado cuando el conjunto de ítems analizado provenga de un componente único, pero también puede ser elevado si el conjunto de ítems obedece a una estructura multidimensional, donde diferentes dimensiones tienen un peso diferencial pero destacable el proceso de respuesta.

En un estudio de simulación Monte Carlo, Green, Lissitz y Mulaik (1977) generaron un test de 10 ítems que medían cinco componentes. Los cinco componentes fueron ortogonales, y cada ítem tuvo la misma carga (.45) en dos de los componentes, aunque ninguno de ellos tuvo sus cargas factoriales en el mismo par de dimensiones. La comunalidad de cada ítem fue de .90. El coeficiente alfa resultante para el test completo fue de .81, lo que podría interpretarse como una evidencia fuerte de la undimensionalidad de este conjunto de ítems, aun siendo ese conjunto de ítems multidimensional.

Cortina (1993) manipuló la intercorrelación promedio (.30, .50 y .70) entre los ítems de tres tests cuyas longitudes fueron 6, 12 y 18 ítems. Entonces, calculó el coeficiente alfa bajo una estructura que varío desde una única dimensión a tres dimensiones. El resultado fue que el coeficiente alfa presentó una fuerte dependencia de la longitud del test, hecho que no siempre es tenido en cuenta por los investigadores, e incrementó su cuantía conforme aumentó la intercorrelación promedio de los ítems. Cuando la estructura del test tuvo más de una dimensión (dos y tres), el coeficiente alfa fue más bajo que en la estructura unidimensional, pero el patrón de resultados fue semejante al obtenido en la solución unifactorial.

Si un AFE (o AFC) sugiere que un test está formado por varios subtests, donde cada uno responde a una dimensión relevante, entonces se considera que las puntuaciones observadas dependen esencialmente de esta dimensión, y se puede proceder a evaluar su fiabilidad. Sin embargo, cuando se construyen tests que evalúan atributos nuevos o poco conocidos, donde la dimensionalidad es desconocida, existe cierta tendencia a obtener la consistencia interna de las puntuaciones observadas independientemente de la dimensionalidad subyacente. En ambientes clínicos, el problema es aún más grave, ya que se obtiene el coeficiente de fiabilidad (generalmente alfa) tanto para los subtests componentes como para la puntuación total obtenida con el test completo. La unidimensionalidad del test se convierte así en un ideal a alcanzar, que sólo se puede entrever entre ítems, y cuya imprecisión saca a la luz otras dimensiones más o menos relevantes y que introducen variación no deseada en las puntuaciones de los tests.

Aunque es un hecho aceptado actualmente en la literatura experimental que el coeficiente alfa no es una evidencia de la unidimensionalidad del test (Cortina, 1993; Hattie, 1985; McDonald, 1999), parece que no está suficientemente estudiado en qué medida la multidimensionalidad puede afectar al coeficiente alfa y a su cuantía. Green et al. (1977) y Cortina (1993) sólo tuvieron en cuenta una estructura ortogonal de los componentes del test. Además, Cortina (1993) demostró la estrecha dependencia del coeficiente alfa de la longitud del test y del aumento de las covarianzas entre los ítems.

Objetivo de la investigación

Para examinar la influencia de la dimensionalidad en el coeficiente alfa hemos optado por utilizar un test de longitud fijada (30 ítems) y simular matrices de respuestas con una y dos dimensiones. Las matrices de una dimensión son la línea base, ya que son la verdadera expresión de la fiabilidad de las puntuaciones, con las que se compararán los coeficientes alfa obtenidos a partir de las matrices bidimensionales. Las dos dimensiones pueden ser ortogonales u oblicuas, teniendo en cuenta también la importancia de la dimensión a través del número de cargas significativas en cada uno, aunque respetando que la primera dimensión siempre obtiene más cargas significativas que la segunda.

Método

Condiciones experimentales

Para cumplir los objetivos arriba señalados, en este estudio de simulación Monte Carlo hemos manipulado las siguientes variables experimentales: formato del ítem, amplitud del intervalo de dificultad en los ítems, porcentaje de ítems en la segunda dimensión, correlación entre los factores (ortogonales vs. oblicuos), y grado de centralidad de la distribución de la habilidad.

Para los ítems se han manipulado dos formatos, uno para ítems dicotómicos (0/1) (tests de ejecución máxima) y otro para ítems politómicos (p.e. de 1 a 5), más conocidos como ítems tipo Likert (tests de ejecución típica). No obstante, el formato de los ítems politómicos de este estudio es ligeramente diferente al tipo Likert, ya que la categoría más baja se situó en 0 y la más elevada en 4.

Los tests, en función de su ámbito de aplicación, también pueden diferir en el intervalo de dificultad de los ítems que los componen. Así, en un test de ejecución máxima, esperamos que la separación de los índices de dificultad de los ítems sea suficientemente elevada para recoger respuestas representativas de todo el intervalo de habilidad en el rasgo evaluado, mientras que en los tests de ejecución típica, la dispersión de las medias de los ítems puede ser algo menor. Tanto en el formato dicotómico como politómico, las máximas discriminaciones entre las personas se producirán cuando las medias de los ítems sean iguales, pero no es deseable que todos los ítems tengan la misma media, ya que esto dará lugar a una partición del grupo en dos, los que saben (conocen) y los que no saben (no conocen) (Crocker y Algina, 1986). Así que es preferible utilizar un rango más o menos amplio para obtener puntuaciones que cubran la mayor parte del rango de habilidad de acuerdo con la distribución normal. Para reproducir ambas situaciones hemos empleado dos variantes. En la primera, las medias de los ítems son aproximadamente iguales, como puede ocurrir en un test de personalidad o actitudes y menos frecuentemente en test de rendimientos o capacidades, por lo que se eligió un intervalo de dificultad estrecho [-.5, .5]. En la segunda, las medias varían ampliamente dentro del intervalo especificado por el formato del ítem (0/1, ítems dicotómicos) ó (p. e., 0 a 4, ítems politómicos), tal como suele ocurrir en los test de rendimientos o capacidades y menos frecuentemente en los tests de personalidad y actitudes, por lo que el intervalo de dificultad seleccionado fue amplio [-2, +2].

El porcentaje de ítems que cargan en cada dimensión también es una característica importante a tener en cuenta, ya que en un AFE, la importancia de los componentes siempre viene determinada por su orden, y se plasma en la cuantía de las cargas factoriales. Así, el número de cargas factoriales significativas en el primer factor suele ser mayor que en el segundo y subsiguientes factores. Para simular esta situación, hemos manipulado tres porcentajes: 10%, 30% y 50% de ítems que cargan en la segunda dimensión relevante. En un test de 30 ítems, estos porcentajes han producido las siguientes razones de la primera dimensión frente a la segunda: 27:3, 21:9 y 15:15.

La cuarta condición experimental también tiene que ver con la dimensionalidad del test. En este caso, se manipuló la relación entre las dos dimensiones: .00 (ortogonal), .10, .40 y .80 (oblicuas), que se corresponden con correlaciones de .32, .63 y .89 respectivamente.

La habilidad media del grupo también fue manipulada con el objeto de comprobar si los resultados podían cambiar significativamente en grupos que no se comparan con la dificultad media del test, situación que produce cierta inestabilidad y es indeseable cuando se calibra un test con la teoría de las puntuaciones verdaderas. Así, un grupo tuvo media 0 y desviación típica 1, mientras que otro grupo tuvo su habilidad media una desviación típica por debajo de la media (-1), manteniendo su desviación típica en 1.

El cruce de las cinco condiciones experimentales (2 tipos de categorías x 2 intervalos de dificultad x 3 porcentajes de ítems en la segunda dimensión x 4 correlaciones entre dimensiones x 2 habilidades medias) produjo 96 casillas. Cada casilla fue replicada en cinco ocasiones.

Para comparar los resultados de las distintas condiciones experimentales se generaron tests unidimensionales en función del tipo de formato, el número de categorías y las dos medias de habilidad. El resultado fueron ocho matrices de respuestas unidimensionales que fueron replicadas en cinco ocasiones, y cuyo coeficiente alfa promedio (en cada condición) se tomó como línea base de comparación de los coeficientes alfa resultantes en las distintas condiciones experimentales.

Simulación de las matrices de datos

Generalmente, en los estudios que emplean puntuaciones observadas, las respuestas de las personas se han generado a partir de matrices de varianza-covarianza o matrices de correlaciones. En este estudio, sin embargo, hemos preferido utilizar la familia de modelos de Rasch (Rasch, 1980; Wright y Masters, 1982), ya que permiten simular con bastante exactitud tests uni- y bidimensionales de acuerdo con una estructura preestablecida. En la familia de modelos de Rasch, la undimensionalidad es una condición necesaria, por lo que cualquier conjunto de ítems que ajuste este modelo se puede afirmar que es esencialmente unidimensional.

Para generar las matrices de respuestas unidimensionales con ítems dicotómicos se utilizó la función logística del modelo de Rasch dicotómico:

$$P_{ni} = \frac{1}{1 + exp(\beta_n - \delta_i)} \tag{3}$$

donde P_{ni} es la probabilidad de que la persona de habilidad β_n acierte el ítem *i*, y δ_i es la dificultad del ítem *i*. Para generar las matrices de respuestas unidimensionales con ítems politómicos se empleo el modelo de crédito parcial (Wright y Masters, 1982):

$$P_{nix} = \frac{exp \sum_{j=1}^{x} (\beta_n - \delta_{ij})}{1 + \sum_{k=1}^{mi} exp \sum_{j=1}^{k} (\beta_n - \delta_{ij})} \quad para\ x = 1, 2, \cdots, m_i \tag{4}$$

donde P_{nix} es la probabilidad de que la persona de habilidad β_n conteste el ítem de dificultad δ_i en la categoría *x*, donde *x* varía entre 1 y m_i categorías, y el número de parámetros de las categorías (umbrales) a estimar es *j-1.* Ya que en nuestro estudio los ítems tienen cinco categorías, el modelo de crédito parcial estima cuatro umbrales para cada ítem.

Todas las matrices de este estudio fueron simuladas utilizando rutinas de generación de respuestas aleatorias con una versión beta del programa ConQuest (Wu, Adams y Wilson, 1998).

Análisis Estadístico

En este estudio, el tipo de formato del ítem (dicotómico vs. politómico) fue utilizado como un medio para representar ítems típicos de los tests psicológicos, pero es un hecho reconocido que el coeficiente de fiabilidad, y por ende, el coeficiente alfa mejora sustancialmente conforme el número de categorías aumenta. Así, una comparación de estas dos condiciones resultaría ciertamente engañosa y podría enmascarar la verdadera tendencia que se pueda observar en los resultados de las condiciones experimentales aquí empleadas. Por ello, hemos optado por analizar los resultados de este estudio de simulación en cada uno de los formatos empleados, independientemente uno de otro utilizando el paquete estadístico SPSS 15.0.

Por otra parte, un estudio inferencial de los resultados obtenidos con el coeficiente alfa en las distintas condiciones experimentales implica que al menos se cumpla aproximadamente el supuesto de distribución normal de la variable dependiente. Es un hecho patente que el coeficiente alfa, que varía entre 0 y 1, no sigue una distribución normal, por lo que hemos optado por una transformación de raíz cúbica (Barchard y Hakstian, 1997, Hakstian y Whalen, 1976) que sigue una distribución aproximadamente normal con media cero y desviación típica uno (Feldt y Charter, 2006), y cuya forma es: $t = (1-\alpha)^{1/3}$. Los límites de esta transformación se encuentran entre t = 1, para un coeficiente alfa de .001, y t = .100 para un coeficiente alfa de .999.

Resultados

La tabla 1 presenta los promedios (y desviaciones típicas) de las transformaciones del coeficiente alfa para la línea base y en las distintas condiciones experimentales manipuladas en este estudio en tests con ítem dicotómicos. Como se aprecia en la tabla, la transformación t del coeficiente alfa para la línea base (tests unidimensional) siempre estuvo por debajo (recuérdese que la transformación t debe interpretarse en sentido inverso al del coeficiente alfa) de la transformación t del coeficiente alfa en las distintas condiciones experimentales. Cuando las dimensiones son ortogonales (r = .00) además, se aprecia cómo la transformación t del coeficiente alfa aumenta conforme se incrementa el porcentaje de ítems de la segunda dimensión; es decir, conforme se hace más importante la presencia de esa segunda dimensión en términos del número de ítems que la explica.

Tabla 1.
Media (Desviación típica) de los coeficientes alfa obtenidos en test con ítems dicotómicos

		N(0,1)		N(-1,1)	
		B1	B2	B1	B2
	Base	.517 (.0107)	.538 (.0150)	.539 (.0145)	.548 (.0432)
10%	.00	.556 (.0139)	.591 (.0159)	.565 (.0146)	.598 (.0112)
	.10	.548 (.0197)	.578 (.0152)	.565 (.0175)	.587 (.0168)
	.40	.529 (.0159)	.559 (.0147)	.547 (.0143)	.574 (.0204)
	.80	.524 (.0153)	.555 (.0173)	.535 (.0105)	.552 (.0231)
30%	.00	.609 (.0194)	.642 (.0097)	.635 (.0136)	.656 (.0111)
	.10	.602 (.0144)	.643 (.0205)	.618 (.0226)	.647 (.0266)
	.40	.568 (.0147)	.602 (.0247)	.583 (.0177)	.610 (.0100)
	.80	.533 (.0065)	.551 (.0066)	.544 (.0195)	.573 (.0153)
50%	.00	.640 (.0072)	.665 (.0145)	.652 (.0293)	.683 (.0216)
	.10	.629 (.0196)	.656 (.0248)	.630 (.0254)	.659 (.0254)
	.40	.583 (.0155)	.602 (.0172)	.603 (.0164)	.619 (.0214)
	.80	.538 (.0155)	.567 (.0140)	.541 (.0252)	.586 (.0073)

El mismo efecto se produce cuando las dimensiones están relacionadas, pero además, se observa que conforme aumenta la correlación entre las dimensiones, la transformación t del coeficiente alfa se acerca más a la transformación t del coeficiente alfa de la línea base. Por ejemplo, cuando el intervalo de dificultad de los ítems es estrecho [-.5, .5], la transformación t del coeficiente alfa para un test unidimensional formado por 30 ítems dicotómicos es .517 (.0107). Cuando aparece una segunda dimensión ortogonal, la transformación *t* del coeficiente alfa se eleva hasta .556 (.0139), y si la segunda dimensión está relacionada con la primera, la transformación *t* del coeficiente alfa se reduce hasta .548 (r = .10), .529 (r = .40) y .524 (r = .80). Si el porcentaje de ítems en la segunda dimensión se eleva hasta un 30%, entonces la transformación *t* del coeficiente alfa, cuando los dos componentes son ortogonales, se aleja aún más de la línea base (.609), y cuando ambas dimensiones están relacionadas, la transformación *t* del coeficiente alfa se acerca más a la línea base en función del aumento de la correlación entre ambas dimensiones. El mismo patrón interpretativo se produce cuando el intervalo de dificultad de los ítems es [-2, 2], aunque en todas las condiciones experimentales, la transformación *t* del coeficiente alfa ha resultado ligeramente por debajo que cuando el intervalo de dificultad fue más estrecho.

Un ANOVA en cuatro sentidos (sesgo de la distribución, amplitud del intervalo de dificultad, porcentaje de ítems en la segunda dimensión y correlación entre las dimensiones) dio como resultado que todos los efectos principales fueron altamente significativos, pero sólo la interacción de porcentaje de ítems en la segunda dimensión con el grado de correlación entre los dos factores fue significativa y tuvo un tamaño del efecto elevado (F (6, 240) = 17.951, p = .000, η^2 = .359). Así, conforme aumenta el porcentaje de ítems de la segunda dimensión del 10% al 50% (Figura 1), la transformación *t* del coeficiente alfa aumenta, pero también dentro de cada porcentaje, si las dos dimensiones están relacionadas, se observa como un aumento de la correlación entre ambas dimensiones produce un aumento también de la transformación *t* del coeficiente alfa. En términos del coeficiente alfa, esta interacción quiere decir que conforme aumenta la importancia de la segunda dimensión, el coeficiente alfa promedio se reduce, pero si las dimensiones son oblicuas, esta reducción del coeficiente alfa se ve suavizada en función del aumento de la correlación entre las dos dimensiones.

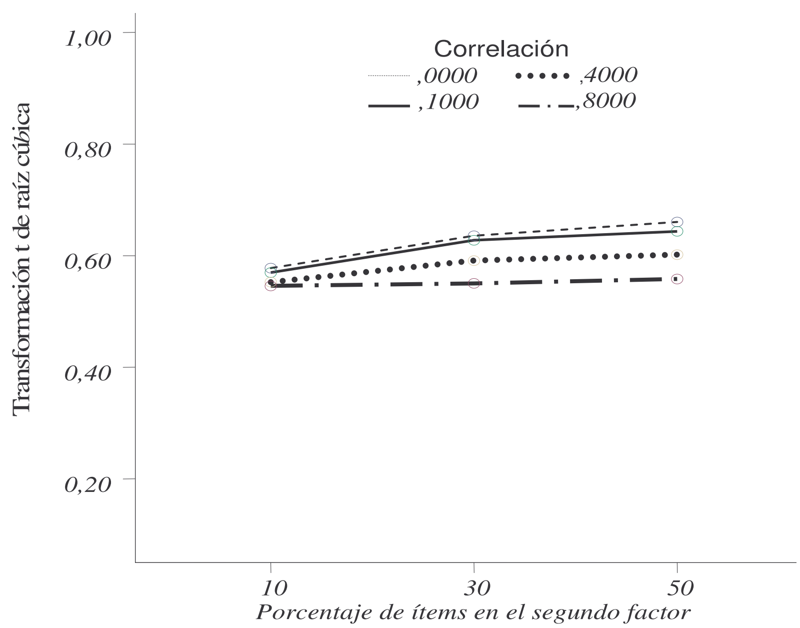

Figura 1: Transformación t del coeficiente alfa según porcentaje de ítems de la segunda dimensión en ítems dicotómicos

La tabla 2 presenta las transformaciones *t* del coeficiente alfa (y desviaciones típicas) para la línea base y las distintas condiciones experimentales manipuladas en este estudio en tests con ítems que siguen un formato politómico. En este caso, dado que el formato de los ítems varía entre 0 y 4, hemos obtenido coeficientes alfa más elevados que en el caso dicotómico. Esto se produce no porque las puntuaciones de los tests con ítems dicotómicos sean menos fiables, sino porque los ítems politómicos incrementan artificialmente la variabilidad muestral, lo que produce que las violaciones del orden de las personas en la muestra se pueda producir con menos frecuencia que con ítems dicotómicos. Así que, por regla general, el mismo test aplicado a una muestra obtendrá un coeficiente de fiabilidad más elevado con ítems politómicos que con ítems dicotómicos, por lo que los coeficientes de fiabilidad no son comparables con los de test formados con ítems dicotómicos.

De nuevo, pero con diferencias mucho menos apreciables, un incremento del porcentaje de ítems en la segunda dimensión (dimensiones ortogonales) produce un incremento de la transformación *t* del coeficiente alfa. Así, si la línea base para el test de 30 ítems con cinco categorías es de .301 (.0061); cuando el porcentaje de ítems en la segunda dimensión es del 10%, la transformación *t* del coeficiente alfa se eleva hasta .345 (.0089), con un 30% hasta .416 (.0073) y con un 50% hasta .439 (.0125). Si las dos dimensiones son oblicuas, entonces el alejamiento de la transformación *t* del coeficiente alfa también aumenta con el incremento de la importancia de la segunda dimensión desde un 10% de ítems hasta un 50%, pero la transformación t del coeficiente alfa cada vez se acerca más a la transformación t del coeficiente alfa de la línea base dentro de cada uno de los porcentajes conforme aumenta la correlación entre las dos dimensiones. Por ejemplo, en un test formado por un grupo de 30 ítems unidimensionales en un intervalo amplio de dificultad [-2, 2], la transformación t del coeficiente alfa (línea base) es .358 (.0078), pero si en el test aparece una segunda dimensión ortogonal (r = .00) con la primera (10%), la transformación t del coeficiente alfa aumenta hasta .402 (.0172), y disminuye paulatinamente a 382 (.0181) cuando la correlación entre ambas dimensiones es .40, y a .366 (.0085) cuando la correlación entre ambas dimensiones es de .80.

Tabla 2.
Media (Desviación típica) de los coeficientes alfa obtenidos en tests con ítems politómicos

		N(0,1)		N(-1,1)	
		B1	B2	B1	B2
	Base	.301 (.0061)	.358 (.0078)	.320 (.0084)	.378 (.0026)
10%	.00	.345 (.0089)	.388 (.0064)	.364 (.0081)	.425 (.1036)
	.10	.341 (.0081)	.402 (.0172)	.356 (.0120)	.413 (.0133)
	.40	.329 (.0093)	.382 (.0181)	.343 (.0101)	.392 (.0217)
	.80	.308 (.0081)	.366 (.0085)	.326 (0114)	.378 (.0161)
30%	.00	.416 (.0073)	.470 (.0165)	.428 (.0055)	.482 (.0108)
	.10	.397 (.0131)	.448 (.0097)	.412 (.0108)	.464 (.0135)
	.40	.364 (.0089)	.407 (.0131)	.381 (.0106)	.432 (.0159)
	.80	.320 (.0071)	.377 (.0150)	.340 (.0140)	.398 (.0106)
50%	.00	.439 (.0125)	.491 (.0104)	.457 (.0109)	.507 (.0166)
	.10	.421 (.0098)	.473 (.0201)	.440 (.0127)	.488 (.0093)
	.40	.377 (.0092)	.428 (.0105)	.398 (.0119)	.441 (.0141)
	.80	.327 (.0094)	.380 (.0105)	.341 (.0119)	.401 (.0227)

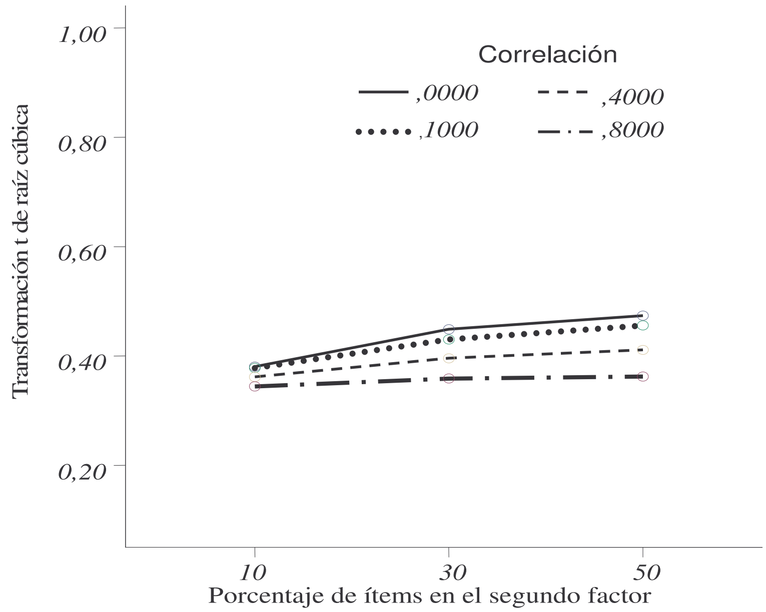

Figura 2: Transformación t del coeficiente alfa según porcentaje de ítems de la segunda dimensión en test politómicos

Un ANOVA en cuatro sentidos dio como resultado que todos los efectos principales fueron altamente significativos, pero sólo la interacción de porcentaje de ítems en la segunda dimensión con el grado de correlación entre las dos dimensiones tuvo un tamaño del efecto elevado (F (6, 240) = 36.705, p = .000, η^2 = .534). Así, conforme aumenta el porcentaje de ítems de la segunda dimensión del 10% al 50% (figura 2), la transformación *t* del coeficiente alfa aumenta, pero también dentro de cada porcentaje, si las dos dimensiones están relacionadas, se observa como un aumento de la correlación entre ambas dimensiones produce un aumento también de la transformación *t* del coeficiente alfa.

DISCUSIÓN

Es un hecho reconocido que el coeficiente alfa no es un indicador de la unidimensionalidad de un test psicológico, pero cómo la multidimensionalidad influye en este coeficiente no parece haber sido abordada con la misma frecuencia. La práctica metodológica en la construcción de tests aconseja que, para darle algún sentido a la puntuación observada cuando el test mide varias dimensiones, el test se debe dividir en tantos subtests como dimensiones esenciales o suficientes se puedan formular, de acuerdo con alguna de las técnicas que se utilizan con frecuencia en este tipo de estudios como el AFE o el AFC.

La práctica psicológica se orienta cada vez más tener en cuenta esta situación y realizar este proceso de división del test en tantos subtests como dimensiones revele la técnica utilizada. Sin embargo, con más frecuencia de la deseada, el investigador ofrece un coeficiente alfa para el total de ítems del test, sin tener en cuenta que el modelo clásico de test asume que la estimación del coeficiente de fiabilidad se debe calcular cuando el test mide esencialmente una variable latente (Gramham, 2006; Miller, 1995). Si el test mide más de una dimensión, entonces este coeficiente de fiabilidad no tiene ningún significado, ya que no sabemos en qué medida contribuyen a la puntuación total las puntuaciones en cada una de las dimensiones relevantes.

Además, como se muestra en este estudio de simulación, el coeficiente de fiabilidad (alfa) se ve influenciado por algunas de las condiciones experimentales más comunes que se suelen presentar en la práctica psicológica. Un Análisis Factorial Exploratorio siempre ofrece una misma estructura interpretativa: el primer componente es más importante (en el cargan más ítems) que el segundo y subsiguientes. En este estudio mostramos como la importancia del segundo componente cuando es ortogonal con el primero reduce drásticamente el coeficiente alfa conforme el número de ítems de la segunda dimensión aumenta. En el caso de que ambas dimensiones sean oblicuas, si la correlación entre ellas es baja, el coeficiente alfa de nuevo se ve disminuido grandemente sobre la estructura unidimensional, pero conforme la correlación entre ambas dimensiones oblicuas aumenta, el coeficiente de fiabilidad tiende al obtenido en la línea base (unidimensional), situación lógica ya que si la correlación entre las dos dimensiones es mayor de .60, prácticamente el AFE está anunciando que sólo una dimensión es la que explica la mayor parte de la covariación entre los datos (Gorsuch, 1983).

Aunque el ANOVA ha encontrado diferencias altamente significativas entre los efectos principales, los tamaños del efecto correspondientes fueron importantes (Cohen, 1988), por lo que no parece que el sesgo de la distribución de las puntuaciones o la amplitud del intervalo de la dificultad de los ítems sean factores influyentes en la cuantía del coeficiente alfa. También los resultados se repiten prácticamente igual en los ítems dicotómicos como en los politómicos, aunque la cuantía de la fiabilidad en tests con ítems politómicos es mayor, por lo que la influencia de la presencia de multidimensionalidad en los ítems se puede generalizar tanto al ámbito de los ítems dicotómicos como el de los ítems politómicos.

Por último, un estudio de simulación tiene un valor limitado ya que es válido sólo para las condiciones experimentales en él empleadas, pero creemos que este estudio de simulación ha mostrado que la presencia de multidimensionalidad en un test psicométrico influye en el coeficiente alfa disminuyéndolo en una cuantía desconocida, aun cuando el investigador considere que esa segunda dimensión es poco relevante. Esta influencia es tanto mayor cuando las dimensiones son ortogonales y se reduce de forma importante hasta desaparecer cuando las dimensiones están altamente relacionadas. En consecuencia, dado que no es posible formular un test estrictamente unidimensional, y en la práctica, las personas siempre necesitarán de más de un atributo o dimensión para contestar los ítems del test, la cuantía del coeficiente alfa se verá disminuida por la presencia de esa segunda dimensión en función del porcentaje de ítems que contiene, y su relación con la dimensión relevante, lo que supone que el coeficiente alfa probablemente será una infraestimación de la fiabilidad.

Referencias

Barchard, K. A. y Hakstian, A. R. (1997). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. Multivariate Behavioral Research, 32, 169-191.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78, 98-104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 15, 297-334.

Crocker, L. y Algina, J. (1986). An introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.

Feldt, L. S. y Charter, R. A. (2006). Averaging internal consistency reliability coefficients. Educational and Psychological Measurement, 66, 215-227.

Green, S. B., Lissitz, R. W. y Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement, 37, 827-838.

Gorsuch, R. L. (1983). Factor analysis, 2 ed. Hillsdale, New Jersey: LEA.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. Educational and Psychological Measurement, 66, 930-944.

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley.

Hambleton, R. K. y Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer-Nijhoff.

Hakstian, A. R. y Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. Psychometrika, 41, 219-231.

Hattie, J. A. (1985). Methodological review: Assessing unidimensionality of test and items. Applied Psychological Measurement, 9, 139-164.

Kaiser, H. (1968). A measure of the average intercorrelation. Educational and Psychological Measurement, 28, 245-247.

Kuder, G. F. y Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2, 151-160.

Lord, F. M. y Novick, M. R. (1968). Statistical theories of mental test scores. Reading Mass.: Addison-Wesley.

McDonald, R. (1999). Test theory: A unified treatment. New Jersey: LEA.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. Structural Equation Modeling, 2, 255-273.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment test. Chicago: University Chicago Press.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. Psychometrika, 52, 589-617.

Suen, H. K. (1990). Principles of test theories. Hillsdale, NJ: LEA.

Wright, B. D. y Masters, G. N. (1982). Rating scale analysis. Chicago: MESA Press.

Wu, M. L., Adams, R. J. y Wilson, M. R. (1998). Acer ConQuest: Generalised item response modelling software. Australian Council for Educational Research.