PEL
Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana
2015, 52(2), 6-33

Pontificia Universidad Católica de Chile
Facultad de Educación

Componentes de la varianza de escalas con una estructura de subescalas empleando dos cálculos de coeficiente α

Components of Variance of Scales with a Subscale Structure Using Two Calculations of Coefficient α

David Andrich

Graduate School of Education, The University of Western Australia

Resumen

En las mediciones sociales, se construyen escalas para medir una sola variable que están, sin embargo, compuestas de subescalas de ítems que miden diferentes aspectos de la variable. Si bien la presencia de subescalas captura la complejidad de una variable y, por lo tanto, aumenta la validez de la escala, técnicamente, la unidimensionalidad se ve comprometida. Como resultado, la presencia de subescalas ha recibido una atención considerable y, más específicamente, ha llevado a la formulación de una estructura bifactorial en la cual todas las subescalas resumen una variable común y, en la que además, los ítems de cada subescala también resumen un aspecto único de dicha subescala. El presente artículo muestra que, con algunos supuestos comunes simplificados sobre una estructura bifactorial, la proporción de dos cálculos del coeficiente α, uno a nivel de ítems y el otro a nivel de subescalas, puede usarse para obtener (a) la proporción de la varianza común verdadera, (b) la proporción de la varianza única verdadera, (c) la proporción de la varianza común verdadera relativa a la suma de las varianzas común y única verdaderas y (d) la correlación combinada entre subescalas corregida para la atenuación debida al error. El artículo sugiere que, atendiendo a que los cálculos son relativamente simples, pueden ser usados para obtener un resumen más completo de las propiedades de una escala que cuenta con subescalas, comparado con lo que es posible obtener con un solo estadístico, como puede ser algún tipo de coeficiente de confiabilidad. Este artículo incluye el ejemplo de un test de aptitud académica, consistente en 100 ítems, que se compone de cuatro subescalas. Un pequeño estudio de simulación muestra que, cuando se satisfacen los supuestos, las estimaciones de las varianzas son estables.

Palabras clave: dimensionalidad, coeficiente alfa, subescalas, estructura bifactorial, componentes de varianza

Correspondencia a:
David Andrich
Graduate School of Education, The University of Western AustraliaM428, 35 Stirling Highway, Crawley, Western Australia, 6009
Correo electrónico: david.andrich@uwa.edu.au
Una versión anterior de este artículo fue presentada en un Taller Internacional de Medición Objetiva, en New York, NY, abril de 2008, y en una Conferencia Internacional sobre Medición de Resultados, en Bethesda, Maryland, septiembre de 2010. La investigación que se reporta en este artículo fue apoyada en parte por una beca de enlace del Consejo de Investigación de Australia. El permiso para usar datos de la Prueba de Aptitud Académica de Australia, que es un test construido por el Consejo Australiano de Investigación Educativa, fue concedido por el antiguo Consejo Curricular de Australia Occidental, hoy la Autoridad de Currículos y Estándares Escolares.

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi: 10.7764/PEL.52.2.2015.2

Abstract

Many scales in social measurement constructed to measure a single variable are nevertheless composed of subscales of items which measure different aspects of the variable. Although the presence of subscales captures the complexity of a variable, and thereby increases the validity of the scale, technically, unidimensionality is compromised. As a result, the presence of subscales has received substantial attention, and in particular, it has led to the formulation of a bifactor structure in which all subscales summarize a common variable and in addition, the items within each subscale also summarize an aspect unique to that subscale. This paper shows that, with some common simplifying assumptions about a bifactor structure, the ratio of two calculations of coefficient α , one at the level of the items, the other at the level of the subscales, can be used to obtain (a) the proportion of true common variance, (b) the proportion of the true unique variance, (c) the proportion of the true common variance relative to the sum of the true common and unique variances, and (d) the summary correlation among subscales immediately corrected for attenuation due to error. The paper suggests that because the calculations are relatively simple, they can be used to provide a more comprehensive summary of the properties of a scale with subscales than is possible with a single statistic such as some form of reliability coefficient. This paper provides an example in which a scholastic aptitude test consisting of 100 items is composed of four subscales. A small simulation study shows that when the assumptions are satisfied, the estimates of the variances are stable.

Keywords: dimensionality, coefficient alpha, subscales, bifactor structure, components of variance

Muchas escalas en psicología, educación, evaluación de salud y medición social, en general, se construyen para medir una única variable. En general, se dice que estas escalas son *unidimensionales*. En todas las teorías de los tests (Lord & Novick, 1968) de escalas unidimensionales, cada persona es caracterizada por un único valor y en la Teoría Clásica de los Tests (TCT) por la puntuación total. Las guías prácticas de construcción de tests incluyen consejos para la construcción de ítems que son más probables de ser estadísticamente independientes (Mehrens & Lehman, 1991) al evaluar la misma variable. Las dos razones principales por las cuales las escalas se componen de más de un ítem estadísticamente independiente son: primero, mientras mayor sea el número de ítems, potencialmente se tendrá un mayor número de puntuaciones totales y, por lo tanto, se tendrá potencialmente una mayor precisión; segundo, mientras mayor sea el potencial número de *aspectos* de la misma variable que se evalúen, mayor será su potencial validez.

Muchas escalas construidas para ser unidimensionales también tienen más de un ítem para evaluar cada uno de los múltiples aspectos de la variable. La escala, entonces, explícitamente tiene *subescalas de ítems*. Aunque por definición todas las subescalas evalúan una variable común, cada una de ellas también evalúa una variable única de su aspecto o rasgo. Como artefacto, técnicamente la escala ya no es unidimensional. Sin embargo, dado que se considera que en conjunto evalúan la variable común de manera más válida que si hubiese solo un ítem por aspecto, las subescalas con múltiples ítems se conservan.

La presencia de subescalas se presta para hipotetizar una simple estructura bifactorial (Holzinger & Swineford, 1937), la cual se ha reconsiderado en contextos relacionados (Chen, West, & Sousa, 2006; Gibbons & Hedecker, 1992; Raykov & Shrout, 2002; Reise, Moore, & Haviland, 2010; Reise, Morizot, & Hays, 2007; Zhang & Stout, 1999) y, por lo tanto, al análisis de los componentes relativos de la varianza representado por los aspectos comunes y únicos. Partiendo de la base de los principios de la Teoría Clásica de los Tests (en adelante TCT) y la estructura bifactorial en la cual se presume la presencia de varianzas únicas homogéneas para las subescalas y de covarianzas homogéneas entre todos los pares de subescalas, junto con presunciones simplificadas comunes de varianzas de error homogéneas y no correlacionadas, el presente artículo muestra que, a partir de los dos cálculos del familiar coeficiente α, uno a nivel de los ítems y otro a nivel de las subescalas, puede calcularse la varianza común verdadera a todas las subescalas, la varianza única verdadera de las subescalas y la varianza de error. Como resultado, pueden calcularse cuatro índices comunes que resultan de interés: (a) la correlación global de los pares de subescalas corregida inmediatamente para la atenuación debida al error; (b) la proporción de la varianza total observada, que es la suma de las varianzas común y única de las subescalas; (c) la proporción de la varianza total, que es la varianza común; y (d) la proporción de la varianza común relativa a la suma de las varianzas común y única.

Bajo ciertas circunstancias, el coeficiente α proporciona una estimación de la confiabilidad de una escala definida en la TCT como la proporción entre la puntuación verdadera a la varianza total, siendo esta última varianza la suma de la varianza verdadera y la de error. La literatura sugiere la existencia de muchos obstáculos para interpretar α como un índice de confiabilidad, por lo que se han sugerido recomendaciones para superarlos (por ejemplo, Cortina, 1993; Green, Lissitz, & Mulaik, 1977; Komaroff, 1997; McDonald, 1978; Rae, 2006: Raykov, 1998; Schmitt, 1996; Sijtsma, 2009; Van Zyl, Neudecker, & Nel, 2000; Zinbarg, Revelle, Yovel, & Li, 2005), desde que Cronbach (1951) elaborara α basándose en Guttman (1945).

Las connotaciones del término *confiabilidad* son evaluativas y positivas que implica que todo es mejor si se tienen mayores valores de la confiabilidad. Es probablemente debido a esta connotación, junto con el intento de encontrar y sugerir las conceptualizaciones ideales de confiabilidad (por ejemplo, Zinbarg et al., 2005), que existe una vasta literatura sobre este tema. El presente artículo no apunta a contribuir a la literatura de este tipo. En lugar de ello, se enfoca en explotar los cálculos simples de α para obtener los componentes de varianza de una escala que cuenta con subescalas definidas a priori cuando se aplican algunos supuestos simplificados a la estructura de las subescalas. Luego, dada esta información, dejo a criterio del investigador o del lector emitir juicios evaluativos de una escala específica dentro del contexto relevante, lo cual podría incluir análisis adicionales. Por lo tanto, este artículo se ocupa solo incidentalmente de la definición de confiabilidad de la TCT, a la cual se hace una referencia somera. Además, este artículo reconoce que la teoría de la generalizabilidad se enfoca en la descomposición de la varianza (Brennan, 1997). Sin embargo, se sugiere que dado que el cálculo de α es aparentemente ubicuo y que los cálculos propuestos son relativamente simples, estos pueden usarse para entregar fácilmente un resumen de los componentes de la varianza de una escala con subescalas.

El artículo formaliza y explota la observación de que α, calculado al nivel de los ítems originales, es mayor que cuando se calcula al nivel de las subescalas (por ejemplo, Marais & Andrich, 2008; Rae, 2006; Smith, 2005; Zenisky, Hambleton, & Sireci, 2002). El resto del artículo se estructura de la siguiente manera: la Sección 2 resume las inferencias del coeficiente α cuando se calcula al nivel de los ítems y al nivel de las subescalas, la Sección 3 muestra los análisis y la interpretación de los componentes de la varianza de un conjunto de datos reales y, la Sección 4 contiene un resumen y la discusión. Dado que el artículo se concentra en la proporción del coeficiente α calculado bajo dos condiciones, con fines de completarlo, la mayoría de las formulaciones matemáticas se presentan en los Apéndices.

Formalización de una estructura de subescala y del coeficiente α

Anticipándome a la formalización de la subescala, con una estructura bifactorial, comenzaré por resumir el cálculo de α donde los supuestos de la TCT (Gulliksen, 1950; Lord & Novick, 1968) se cumplen y no existe estructura de subescalas. La puntuación observada de una persona en un ítem se resuelve en primer lugar definiendo sus componentes verdaderos y de error; que se relaciona a su vez con la resolución similar que se hace habitualmente de la puntuación total en una escala.

Varianzas TCT en el caso estándar

Sea la puntuación observada de la persona n en el ítem i, $i = 1,2,...,I$, x_{ni}, $x_{ni} \in \{0,1,2,..m_i\}$, donde m_i es la puntuación máxima del ítem i. Puede resolverse x_{ni} de acuerdo a

$$x_{ni} = \tau_n + \varepsilon_{ni} \qquad (1)$$

donde τ_n es el valor en la variable τ común a las respuestas de la persona n a todos los ítems y donde ε_{ni} es el componente del error de esta respuesta al ítem i. En el presente artículo τ_n se denomina *puntuación verdadera común del ítem* (aunque el uso del término "común" es estrictamente redundante en este caso, se emplea para distinguirlo de puntuación "única" en el contexto de una estructura de subescalas). Si bien al nivel de una persona n los valores son fijos, para una población relevante de personas τ y ε_i se toman como variables aleatorias. Si ε_i no se correlaciona con los valores de la persona y tiene una distribución homogénea entre los ítems y normal con media 0 y varianza σ_ε^2: $N(0, \sigma_\varepsilon^2)$. Aunque no es estrictamente necesario, por razones de conveniencia, la distribución de personas también será normal. Luego

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2),\ \tau \sim N(\mu, \sigma_\tau^2),\ COV[\tau, \varepsilon_i] = 0, \qquad (2)$$

y de las ecuaciones (1) y (2)

$$V[x_i] = V[\tau] + V[\varepsilon], \qquad (3)$$

esto es $\sigma_i^2 = \sigma_\tau^2 + \sigma_\varepsilon^2$, $\qquad (4)$

donde $\sigma_i^2 = V[x_i]$; $\sigma_\tau^2 = V[\tau]$ y $\sigma_\varepsilon^2 = V[\varepsilon]$, en las cuales, debido a la homogeneidad, no hay necesidad de usar el subíndice i. Para facilitar la exposición, la notación de operadores $V[.]$, $COV[.]$ se usa en las derivaciones y la notación σ^2 para los valores resultantes.

Sea $y_n = \sum_{i=1}^{I} x_{ni}$ la puntuación total de la persona n en la escala compuesta por I ítems, obteniéndose

$$\begin{aligned} y_n &= \sum_{i=1}^{I} x_{ni} = \sum_{i=1}^{I} (\tau_n + \varepsilon_{ni}) = \sum_{i=1}^{I} \tau_n + \sum_{i=1}^{I} \varepsilon_{ni} \\ &= I\tau_n + \sum_{i=1}^{I} \varepsilon_{ni}. \end{aligned} \tag{5}$$

Además, sea

$$t_n = I\tau_n; \; e_n = \sum_{i=1}^{I} \varepsilon_{ni}. \tag{6}$$

Entonces, a partir de (5) y (6),

$$y_n = t_n + e_n, \tag{7}$$

donde t_n es la puntuación verdadera de la TCT y e_n es el error en la escala completa para la persona n. Nos referimos a t_n como *puntuación verdadera de la escala común* para distinguirla de la puntuación común verdadera de los ítems τ_n. Sobre la base de la ecuación (7),

$$\begin{aligned} &V[y] = V[t] + V[e], \\ &\sigma_y^2 = \sigma_t^2 + \sigma_e^2 . \end{aligned} \tag{8}$$

Las ecuaciones (7) y (8) caracterizan la TCT.

Sobre la base de las presunciones de que los errores son homogéneos entre ítems y que $COV[\varepsilon_i, \varepsilon_j] = 0$ para todo i, j, de acuerdo a la ecuación (6),

$$\begin{aligned} &\sigma_t^2 = V[I\tau] = I^2 V[\tau] = I^2 \sigma_\tau^2, \\ &\sigma_e^2 = V[e] = IV[\varepsilon] = I\sigma_\varepsilon^2. \end{aligned} \tag{9}$$

Será conveniente usar la ecuación (9) en las derivaciones e interpretaciones de este artículo. Una observación inmediata basada en la ecuación (9) es que aunque la varianza de los errores aumenta solo linealmente como una función del número de ítems, la varianza de las puntuaciones verdaderas de la escala común aumenta cuadráticamente, lo cual implica que, a medida que crece el número de ítems, la varianza se incrementa en relación con la varianza de error.

La confiabilidad tradicional (teórica), denotada por ρ_{yy}, se define aquí por esta razón de varianzas:

$$\rho_{yy} = \frac{\sigma_t^2}{\sigma_y^2} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_e^2}, \tag{10}$$

la cual es claramente la proporción de la varianza de la puntuación verdadera de la escala común, σ_t^2, en relación con la varianza de la puntuación total observada σ_y^2.

Se asume que las variables τ_n, ε_{ni}, t_n, e_n son continuas; por lo tanto, la implicación es que x_{ni} e y_n también son continuas. Sin embargo, puesto que x_{ni} generalmente solo toma valores enteros, generalmente 0 y 1 por tratarse de ítems puntuados dicotómicamente, y_n también adquiere valores enteros discretos por lo que la continuidad implícita en la ecuación (7) es violada. Discuto sobre esta violación y sobre una implicación mayor más adelante, pero por ahora asumo que, como se hace tradicionalmente, la aproximación de puntuaciones discretas para puntuaciones continuas es satisfactoria.

Una revisión del cálculo del coeficiente α

La deducción de la ecuación para calcular α, se presenta en extenso en el Apéndice A, que comienza con la resolución de la respuesta a cada ítem de acuerdo a la ecuación (1). Sobre la base del Apéndice A y Cronbach (1951),

$$\alpha = \frac{I}{I-1}\frac{V[\sum_{i=1}^{I} x_{ni}] - \sum_{i=1}^{I} V[x_i]}{V[y]} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2 / I} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_e^2}. \tag{11}$$

La ecuación (11) muestra que cuando los supuestos de la TCT se cumplen a cabalidad, α proporciona un cálculo de la confiabilidad de la TCT; de lo contrario, es un aproximación por defecto (McDonald, 1978). La ecuación (11) también muestra cómo el valor de $\sigma_e^2 = \sigma_\varepsilon^2 / I$ disminuye y, por lo tanto, α aumenta a medida que aumenta el número de ítems. En este artículo, identifico al coeficiente α con la *estructura del cálculo* del miembro izquierdo de la ecuación (11), y considero las consecuencias de este cálculo en presencia de subescalas.

Empleo la notación $V[y]$ para el denominador en los cálculos de α pero para enfatizar la estructura en la ecuación (11), uso $V[\sum_{i=1}^{I} x_i]$ en vez $V[y]$ en el numerador.

Formalización de las subescalas de una escala

La estructura bifactorial se formaliza a nivel del ítem, lo que requiere una calificación de la ecuación (1). De acuerdo con esto, al resolver x_{ni}, se obtiene

$$x_{ni} = \tau_n + c_i \vartheta_{ni} + \varepsilon_{ni}, \tag{12}$$

donde τ_n mantiene el mismo significado que en la ecuación (1), ϑ_{ni} es el valor en único aspecto de la variable de la persona n para el ítem i, c_i es el peso o ponderación del ítem i y ε_{ni} es nuevamente el componente del error en la respuesta al ítem i de la persona *n*. Si cada ítem evalúa un aspecto único no correlacionado con el aspecto único de cualquier otro ítem, es decir, si no existe estructura de subescalas, entonces el aspecto único de cada ítem es absorbido en el error, $c_i\vartheta_{ni} + \varepsilon_{ni} \rightarrow \varepsilon_{ni}$, lo cual conduce a la ecuación (1). Sin embargo, con una estructura de subescalas, y asumiendo que $c_i = c_s$ es el mismo valor para todos los ítems dentro de cada subescala, la varianza única de las subescalas puede separarse de la varianza de error. El resultado de este artículo es que, suponiendo además que $c_s = c$ para todos los ítems entre todas las subescalas, c puede estimarse a partir de dos cálculos de α. Queda claro que especificar $c_i = c_s = c \ \forall i,s$ es una fuerte suposición simplificadora. Sin embargo, es la estimación de c bajo tal suposición simplificadora como ésta lo que permite obtener una panorámica rápida de las aproximaciones de los componentes de la varianza entre las subescalas.

En primera instancia, y para efectos de exposición, se mantendrá el subíndice s en c_s. Entonces, sea

$$\beta_{ns} = \tau_n + c_s\vartheta_{ns}, \tag{13}$$

$c_s \geq 0$, donde τ_n es nuevamente la puntuación común verdadera de los ítems para la persona n entre subescalas y es la misma variable que en la ecuación (1), y ϑ_{ns} es la puntuación única del ítem para todos los ítems de la subescala s. Por razones de conveniencia, y puesto que no tiene término de error, β_{ns} se denomina *puntuación verdadera única de las subescalas*. De acuerdo a la estructura bifactorial $COV[\vartheta_s,\tau] = COV[\vartheta_s,\vartheta_w] = 0$ para todas las subescalas s, w, con lo que se obtiene

$$\begin{aligned} V[\beta_s] &= V[\tau] + c_s^2 V[\vartheta_s], \\ \sigma_{\beta s}^2 &= \sigma_\tau^2 + c_s^2\sigma_{\vartheta s}^2. \end{aligned} \tag{14}$$

Nuevamente, aunque no es estrictamente necesario, distribúyase ϑ_s normalmente en la misma población, $\tau_s \sim N(\mu_s, \sigma_{\vartheta s}^2)$. Más aún, sin perder generalidad y para fines de identificación, sea

$$\sigma_\tau^2 = \sigma_{\vartheta s}^2 \tag{15}$$

para todos los ítems dentro de todas las subescalas *s*. La ecuación (15) convenientemente hace que el valor c_s^2 sea la proporción de la varianza única de la subescala $\sigma_{\vartheta s}^2$ relativa a la varianza de la puntuación verdadera común de los ítems σ_τ^2: $c_s^2 = (\sigma_{\beta s}^2 - \sigma_\tau^2)/\sigma_{\vartheta s}^2 = (\sigma_{\beta s}^2 - \sigma_\tau^2)/\sigma_\tau^2$. Además, la ecuación (15) implica que las varianzas de los ítems dentro y entre subescalas son homogéneas.

Claramente, si $c_s \neq 0$, que es el caso empleado en este artículo, se tiene que al tomar todos los ítems de una escala en conjunto se viola la unidimensionalidad. En las deducciones de este trabajo, se mantienen los términos $V[\tau]$ y $V[\vartheta_s]$ con fines de claridad en la exposición, pero para simplificar las expresiones, el equivalente numérico, $\sigma_\tau^2 = \sigma_{\vartheta s}^2$, de la ecuación (15) se aplica en las expresiones finales de las derivaciones. Por la misma razón, el subíndice s en c_s se mantiene para aplicar $c_s = c$ en las expresiones finales (incidentalmente, puede tomarse $c_s < 0$, pero dado que sus efectos aparecen como el cuadrado c_s^2, $c_s \geq 0$ se impone por conveniencia, implicando que las subescalas no están correlacionadas de manera negativa.)

El Apéndice B presenta los cálculos de los componentes de la varianza que resultan de las ecuaciones (12) y (13). Para simplificar la exposición, supóngase primero que el número de ítems K es el mismo para todas las subescalas S; luego, $I = SK$ es el número total de ítems. Demuestro que este supuesto puede relajarse.

Si $I = SK$, entonces de acuerdo a la ecuación (9),

$$\sigma_t^2 = I^2\sigma_\tau^2 = S^2K^2\sigma_\tau^2,$$
$$\sigma_e^2 = I\sigma_\varepsilon^2 = SK\sigma_\varepsilon^2. \tag{16}$$

La correlación entre ítems de diferentes subescalas

El Apéndice B muestra que ρ_{sw}, la correlación entre las puntuaciones verdaderas de las subescalas de cualquier par de subescalas s y w es

$$\rho_{su} = \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + c_s^2\sigma_{\vartheta s}^2}\sqrt{\sigma_\tau^2 + c_w^2\sigma_{\vartheta w}^2}}. \tag{17}$$

Aplicando $\sigma_\tau^2 = \sigma_{\vartheta s}^2 = \sigma_{\vartheta w}^2; c_s = c_w = c, \forall\, s, w$, se obtiene

$$\rho_{sw} = \frac{1}{1+c^2}, \ \forall\, s, w. \tag{18}$$

Claramente, mientras mayor sea el valor de c, menor será la correlación entre dos subescalas de las diferentes subescalas, si $c = 0$, la correlación es 1, que mantiene la unidimensionalidad, y no existe una estructura de subescalas efectiva. En el caso de solo dos subescalas, ρ_{sw} es idéntico, conceptualmente y en cuanto a su valor, a la correlación entre dos subescalas corregidas por atenuación debido al error, es decir, $\rho_{sw} = 1/(1+c^2) = r_{sw}/\sqrt{r_{ss}}\sqrt{r_{ww}}$ donde r_{sw} es una correlación observada entre las subescalas y r_{ss}, r_{ww} son estimaciones de la confiabilidad de cada subescala. Un único valor de c a través de todos los pares de subescalas hace que ρ_{sw} sea una generalización de la correlación entre dos escalas, corregida para el error, y es un resultado relevante en sí mismo para los propósitos de este artículo. Además, en este artículo, y para contrastarlo con una correlación observada o manifiesta, ρ_{sw} se denomina correlación *latente* entre subescalas.

Representación de las subescalas en el cálculo de α

Para dar cuenta de la estructura de las subescalas, cada subescala $s, s = 1,2,\ldots,S$ toma el papel de un ítem cuya puntuación es la suma de las puntuaciones de los ítems originales en esa subescala. Todos los supuestos de la TCT presentadas más arriba se mantienen al nivel de las subescalas. Nuevamente, la formalización comienza con la resolución de la puntuación de un ítem.

Entonces, resolviendo la puntuación observada x_{nis} de una persona *n* en un ítem *i* de una subescala *s* de acuerdo a

$$x_{nis} = \beta_{ns} + \varepsilon_{nis} = \tau_n + c_s \vartheta_{ns} + \varepsilon_{nis}, \tag{19}$$

donde ε_{nis} es el componente del error del ítem *i* de la subescala *s* por persona y $COV[\tau_s, \varepsilon_{is}] = 0$. Se asume que $\varepsilon_{is} \sim N(0, \sigma_\varepsilon^2)$ es homogéneo tanto entre ítems dentro de una subescala como entre ítems de diferentes subescalas, de modo que σ_ε^2 no lleva los subíndice s ni i. Aunque tiene la misma distribución para todos los ítems, para fines de exposición, mantengo los subíndices i y s en ε_{is}.

Con K ítems por subescala, sea

$$y_{ns} = \sum_{i=1}^{K} x_{nis}. \tag{20}$$

la puntuación observada de la persona n en la subescala S.

El mecanismo para considerar la estructura de las subescalas consiste en tomar la suma de las puntuaciones de cada persona dentro de cada subescala y tratar las puntuaciones resultantes de las subescalas como las unidades de análisis. Luego, la puntuación total de una persona *n* en la escala es

$$y_n = \sum_{s=1}^{S} y_{ns} = \sum_{s=1}^{S} \sum_{i=1}^{K} x_{nis}. \tag{21}$$

Por lo tanto, emplear los resultados en las ecuaciones (A2.4) – (A2.7) del Apéndice B da

$$\begin{aligned} V[y] &= S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2; \; (c_s^2 = c^2; \; \sigma_{\vartheta s}^2 = \sigma_\vartheta^2, \; \forall s) \\ &= \sigma_t^2 + \sigma_u^2 + \sigma_e^2, \end{aligned} \tag{22}$$

donde

$$\sigma_u^2 = SK^2c^2\sigma_\vartheta^2 \tag{23}$$

es la suma de las varianzas únicas en todas las subescalas. Para el resto de este artículo, σ_u^2 y σ_t^2 se denominarán como la varianza única y común respectivamente, entendiéndose que estas son varianzas "verdaderas" en el sentido de que no son varianzas de error. Téngase en cuenta según la ecuación (22) que, a pesar de que cada subescala adicional añade una varianza única, $\sigma_u^2 = SK^2c^2\sigma_\vartheta^2$ que solo interviene linealmente en función de S, cada subescala adicional añade a la varianza común $\sigma_t^2 = S^2K^2\sigma_\tau^2$ cuadráticamente.

Al considerar una estructura de subescalas cuando $c = 0$ y $c \neq 0$ se obtienen diferentes valores para α. Las ecuaciones se derivan en el Apéndice C y se resumen en la Tabla 1, que contiene los numeradores para cuatro cálculos de α. Sus denominadores son idénticos y se muestran en la última línea de la Tabla 1.

La notación de cada cálculo de α también se muestra en la Tabla 1. La primera columna en la primera fila, $c = 0,\ \rho_{sw} = 1,\ \forall s, w$ (se satisface la unidimensionalidad) no considera las subescalas, es el caso estándar, y se anota simplemente como α. La segunda columna de la primera fila es nuevamente $c = 0$, que puede ser un valor empírico, pero la fórmula sí considera una estructura de subescalas y se anota como α_0. La tercera columna muestra la relación entre α y α_0. En la primera columna de la segunda fila, $c > 0$, pero la fórmula no considera la estructura de las subescalas y se escribe α_c. En la segunda columna de la segunda fila, $c > 0, \forall s$, la fórmula sí considera la estructura de las subescalas y se escribe α_c. La tercera columna muestra la relación entre α_c y α_S.

Tabla 1
Condiciones de cálculo para α y sus valores

	Sin considerar la estructura de las subescalas	Considerando la estructura de las subescalas	Efecto sobre α
Caso estándar: $c_S = 0$, $\rho_{sw} = 1$	$\alpha = \dfrac{\sigma_t^2}{\sigma_y^2}$	$\alpha_0 = \dfrac{\sigma_t^2}{\sigma_y^2}$	$\alpha = \alpha_0$.
$c_S > 0$, $\rho_{sw} < 1$	$\alpha_c = \dfrac{\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1)}{\sigma_y^2}$	$\alpha_S = \dfrac{\sigma_t^2}{\sigma_y^2}$	$\alpha_c > \alpha_S$
$\sigma_y^2 = \sigma_t^2 + \sigma_u^2 + \sigma_e^2$			

La primera fila de la Tabla 1 muestra que si $c = 0$ (existe una estructura de subescalas pero sin valores únicos para las subescalas) entonces α tiene el mismo valor sin importar si se considera o no una estructura de subescalas en su cálculo. En los datos, puede existir una ligera diferencia entre los valores de los dos cálculos debido a diferencias al azar entre ítems en algunas subescalas, pero estadísticamente debe ser 0. Entonces, el hecho de que α calculado en las dos formas diferentes produce efectivamente el mismo valor es evidencia de que la unidimensionalidad estadística y empírica de la escala completa prevalece, a pesar de tratarse de una estructura en la cual las subescalas potencialmente evalúan diferentes aspectos.

La segunda fila de la Tabla 1 muestra que si $c > 0$ (aspectos únicos de las subescalas), entonces $\alpha_c > \alpha_S$. Esta desigualdad explica la conocida observación mencionada anteriormente, de que cuando existe una estructura de subescalas, α calculado a nivel de ítems es mayor que cuando las puntuaciones dentro de las subescalas se suman y α es calculado a nivel de subescalas. El mayor valor de α_C surge de la adición de un factor de σ_u^2, la varianza única de las subescalas, en el numerador. En el caso en el cual la estructura de las subescalas se considera, esta varianza única de la subescala efectivamente se suma con la varianza de error, dejando solo una expresión que incluye la varianza verdadera de los ítems σ_t^2 en el numerador. Ya que las expresiones de la fila superior no incluyen c $(c=0)$ ni en el numerador ni en el denominador, los valores $\alpha = \alpha_0 = \sigma_t^2 / \sigma_y^2$ son idénticos, a pesar de que la expresión en la segunda celda de la fila inferior toma la misma forma $\alpha_S = \sigma_t^2 / \sigma_y^2$, porque el denominador de la fila inferior sí incluye c, $\alpha = \alpha_0 \neq \alpha_S$.

El resultado $\alpha_c > \alpha_S$, $c > 0$ también deja claro por qué el cálculo estándar de α (α_c) no indica el grado de unidimensionalidad de una escala, lo que constituye uno de los obstáculos en la interpretación de α. Si los valores calculados muestran $\alpha_c > \alpha_S$, entonces esta desigualdad confirma la varianza única de las subescalas. Ahora es posible explotar la diferencia entre α_c y α_S.

Recuperación de c y ρ_{sw}

Al reordernar α_c / α_S se obtiene una estimación de c. Si se aplica la ecuación (23) se obtiene

$$\begin{aligned}
\alpha_c / \alpha_S &= \frac{\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1)/V[y]}{\sigma_t^2 / V[y]} \\
&= \frac{\sigma_t^2 + SK^2c^2\sigma_\vartheta^2 S(K-1)/(SK-1)}{\sigma_t^2} \\
&= \frac{\sigma_t^2 + S^2K^2c^2\sigma_\vartheta^2(K-1)/(SK-1)}{\sigma_t^2} \\
&= \frac{\sigma_t^2 + \sigma_t^2 c^2(K-1)/(SK-1)}{\sigma_t^2} \\
&= 1 + c^2(K-1)/(SK-1),
\end{aligned} \tag{24}$$

lo que da como resultado

$$\alpha_c / \alpha_S - 1 = c^2(K-1)/(SK-1), \tag{25}$$

$$c^2 = \left(\frac{SK-1}{K-1}\right)\left(\frac{\alpha_c}{\alpha_S} - 1\right) = \left(\frac{S - 1/K}{1 - 1/K}\right)\left(\frac{\alpha_c}{\alpha_S} - 1\right). \tag{26}$$

De c^2, $\rho_{sw}, \forall s, w$, se obtiene directamente de la ecuación (18).

Como se evidencia en la Tabla 1, si $c > 0$, entonces $\alpha_c > \alpha_S$. Por lo tanto, si $S > 1$, y debido que $SK - 1 > K - 1$, la ecuación (26) implica que si $\alpha_c > \alpha_S$, entonces $c^2 > 0$, como se esperaba. Claramente, si $S = 1$, entonces α_S no puede ser calculado, pero por supuesto, en ese caso no hay necesidad de intentar hacerlo. Como se indicó anteriormente, sin pérdida de generalidad, tómese $c = +\sqrt{c^2}$.

El requisito de que cada subescala tenga el mismo número de ítems puede relajarse. Sea k_s el número de ítems en la subescala s. Luego, puede mostrarse que

$$c^2 = \frac{\left(\frac{\alpha_c}{\alpha_S}\right)\left(\frac{S}{S-1}\right)\left(\frac{(\sum_{s=1}^{S} k_s)-1}{(\sum_{s=1}^{S} k_s)}\right)\left(\sum_{s=1}^{S}\underbrace{\sum_{t=1}^{S} k_s k_t}_{s\neq t}\right) - \left(\sum_{s=1}^{S}\underbrace{\sum_{t=1}^{S} k_s k_t}_{s\neq t}\right) - \left(\sum_{s=1}^{S} k_s(k_s-1)\right)}{\left(\sum_{s=1}^{S} k_s(k_s-1)\right)} . \quad (27)$$

Nótese que si $k_s \neq K$ para alguna s, $s = 1,2,\ldots,S$, entonces incluso si $c = 0$, $\alpha_0 < \alpha$. Así, con unidimensionalidad, cuando el número de ítems por subescala es diferente, el valor de α se reduce cuando los ítems se forman en subescalas y se analizan como ítems de orden superior. Sin embargo, ya que con una estructura de subescalas α_S se reduce aún más, la ecuación (27) se mantiene correcta. Números muy distintos de ítems o de puntuaciones por subescala, harán escalas con un número sustancialmente mayor de ítems o de puntos. Por lo tanto, aunque no es necesario tener exactamente el mismo número de ítems o puntos por subescala, para poder interpretar c^2 con confianza presumiéndose homogeneidad de las varianzas de las subescalas, es necesario que el número de ítems o puntos por subescala sea muy similar y de un tamaño razonablemente grande; por ejemplo, unos 25 ítems dicotómicos por subescala o un número equivalente de puntos (por ejemplo, 12 ítems politómicos con una puntuación máxima de 2). En el ejemplo ilustrativo, el número de ítems dicotómicos por subescala va de 23 a 27 y las fórmulas parecen funcionar con precisión.

Ahora paso a detallar la interpretación de los componentes de varianza dado el valor de c^2.

Simplificación de α_c

Sobre la base de α_c en la Tabla 1 y expandiendo $V[y]$,

$$\alpha_c = \frac{\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1)}{\sigma_t^2 + \sigma_u^2 + \sigma_e^2} . \quad (28)$$

Tomando el límite de $(K-1)/(SK-1)$ con un número creciente de ítems K , se obtiene

$$\lim_{K\to\infty} [K-1]/[SK-1] = \lim_{K\to\infty} [(1-1/K)/(S-1/K)] = 1/S, \quad (29)$$

donde $[K-1]/[SK-1] < 1/S$, $\forall\ S > 1$, $K > 1$.

Sustituyendo $1/S$ por $[K-1]/[SK-1]$ en la ecuación (28) resulta

$$\alpha_c < \frac{\sigma_t^2 + \sigma_u^2}{\sigma_t^2 + \sigma_u^2 + \sigma_e^2} . \quad (30)$$

La ecuación (30) muestra que α_c es el límite inferior de la proporción de la suma de las varianzas común y única en relación a la suma de las varianzas común, única y de error (es decir, relativo a la varianza total). Para todos los efectos prácticos, es esta proporción. Por ejemplo, cuando $K=15,\ S=2,\ (1-1/K)/(S-1/K)=0.483\approx 0.5=1/S$. Por lo tanto, puede escribirse $\alpha_c \cong (\sigma_t^2+\sigma_u^2)/(\sigma_t^2+\sigma_u^2+\sigma_e^2)$.

Una interpretación de la razón α_S/α_c

Denótese la razón α_S/α_c mediante α_A. Luego,

$$\alpha_A=\frac{\alpha_S}{\alpha_c}=\frac{\sigma_t^2/\sigma_y^2}{\sigma_t^2+\sigma_u^2 S(K-1)/(SK-1)/\sigma_y^2}$$
$$=\frac{\sigma_t^2}{\sigma_t^2+\sigma_u^2 S(K-1)/(SK-1)}. \tag{31}$$

.

Invocando el límite de la ecuación (29) se obtiene

$$\alpha_A=\frac{\alpha_S}{\alpha_c}>\frac{\sigma_t^2}{\sigma_t^2+\sigma_u^2}. \tag{32}$$

Para efectos prácticos y con un número razonablemente grande de ítems por subescala, $\alpha_A \cong \sigma_t^2/(\sigma_t^2+\sigma_u^2)$. El numerador α_A es claramente la varianza común y el denominador es la suma de las varianzas común y única, lo que hace que α_A sea efectivamente el límite superior de la proporción de la varianza común relativo a las varianzas común y única, es decir, la proporción de la varianza de no error.

Descripción de una escala con cuatro índices

Como ya se ha indicado, existen muchas advertencias en la literatura sobre interpretar cualquier cálculo de α como un índice de confiabilidad. Estas advertencias parecen surgir del uso de un valor de α, lo que puede inducir a error de distintas formas en varias situaciones. Se sugiere, por lo tanto, que una de las complicaciones de reportar e interpretar α en general y, en particular, en presencia de subescalas, es esperar que un valor único sea suficiente para resumir las propiedades de una escala, y en especial algún tipo de confiabilidad de esta. Este grado de simplificación no es posible. Por esta razón, para entregar un resumen completo de las propiedades de una escala desde la perspectiva de los componentes de la varianza, se sugiere que se reporten los cuatro índices derivados anteriormente, los dos cálculos de α, la proporción de la varianza de no error que corresponde a varianza común y la correlación latente entre subescalas.

En primer lugar, α_S, en donde la varianza única de las subescalas es absorbida por la varianza de error, es el más similar a α cuando se cumplen los supuestos de la TCT, y puede reportarse e interpretarse productivamente como la proporción de la varianza total que es varianza común y verdadera. En segundo lugar, α_A, que indica la proporción de varianza común relativa a la suma de las varianzas común y única (e independiente de la varianza de error), puede interpretarse de forma útil como indicador del grado de multidimensionalidad de la escala. También puede interpretarse como un valor de α corregido por atenuación debido a errores aleatorios en las respuestas a los ítems. En tercer lugar, α_c, el valor inflado por la varianza única y el que usualmente se calcula, también es informativo en el contexto de múltiples índices como la proporción entre las varianzas común y única. Reportar e interpretar los tres valores, α_S, α_A, y α_c, junto con el valor de la correlación latente global ρ entre subescalas, probablemente reduzca las malas interpretaciones respecto de un valor único de α calculado en presencia de subescalas. En el ejemplo ilustrativo se muestra un reporte e interpretación de esto.

Propiedades de la estimación de c

En las derivaciones, las estimaciones no se distinguieron de los valores de los parámetros, ya que se entiende que los cálculos de α basados en datos entregan estimaciones. Por lo tanto, el cálculo de c basado en la ecuación (25) o (26), que es función de la razón de los dos cálculos de α, también es una estimación. En este artículo, no tenemos oportunidad de considerar las propiedades muestrales de la estimación de c. Sin embargo, presento algunas simulaciones que ilustran el uso de un conjunto de datos reales.

Antes de concluir esta sección, vuelvo al hecho de que la derivación de α está basada en el supuesto de variables continuas a pesar de que las respuestas a los ítems de una escala generalmente se puntúan con números enteros y que, como resultado, la puntuación total es un entero. La limitación clave que surge de esta característica es que las puntuaciones totales deben tener suficiente precisión para aproximarse suficientemente a la continuidad y así permitir que funcionen las fórmulas. Parece ser que una puntuación máxima en una escala de al menos 25, con una distribución de personas que vaya desde al menos tres hasta un máximo de 22, entrega una precisión suficiente para recuperar los índices derivados en este artículo.

Una advertencia importante es que la distribución no debe sesgarse artificialmente teniendo a muchas personas con una puntuación máxima y muchas otras cercanas al máximo, o cercanas a una puntuación 0. En otras palabras, es importante que los ítems y las personas no estén tan mal alineados como para generar un fuerte efecto de “piso y techo”, lo que puede ocurrir, por ejemplo, si se diseña una escala para una población normal y se administra a una población clínica, o viceversa. Sin embargo, en general, esta es una limitación importante en la aplicación de la TCT y en la interpretación de coeficientes de confiabilidad, sin importar cómo se calculen, puesto que, con la presencia de efectos de “piso y techo”, las intercorrelaciones entre ítems se inflan tanto como cualquier otro coeficiente de confiabilidad.

Análisis de datos

Esta sección ilustra el análisis de un conjunto de datos reales. Aunque este artículo no considera las propiedades estadísticas de la estimación de la variable c en las derivaciones, para indicar la estabilidad de la estimación con el ejemplo, se simula un conjunto de datos sobre la base de una distribución normal, asignándose los valores de las medias, las desviaciones estándar, los números de ítems por subescala, el rango de dificultad de los ítems, y el valor de c según los valores obtenidos de los datos reales. Luego, para comprobar la estabilidad de la estimación de c, se generan diez réplicas y se reporta la media y la desviación estándar de las estimaciones de c.

Simulación de respuestas discretas

Para hacer un paralelo con las respuestas discretas en el ejemplo real, se generaron respuestas simuladas de acuerdo al modelo dicotómico de Rasch como parte de la clase general de modelos de Rasch (Andersen, 1977; Andrich, 1978; Rasch, 1961; Wright & Masters, 1982). Las presunciones centrales de la TCT y de los modelos Rasch, es decir, independencia local de las respuestas, igual discriminación entre ítems y que la puntuación total proporciona toda la información sobre el perfil de respuestas de una persona, son idénticas. Una diferencia importante entre la TCT y los modelos de Rasch es la especificación de los parámetros de los ítems (las dificultades para la evaluación del desempeño). En las simulaciones, estos parámetros se hicieron de modo equivalente a los del análisis del modelo dicotómico de Rasch de los datos reales con el mismo número de ítems en cada una de las escalas como en los datos reales. Se realizaron diez réplicas, manteniendo inalterados todos los parámetros pero con una semilla diferente para generar los números aleatorios, y se reportó la media y la desviación estándar de las estimaciones relevantes. Subrayo el hecho de que estas simulaciones solo buscan ser ilustrativas y no exhaustivas. No hubo sesgo artificial ni en los datos reales ni en los simulados.

Debido a que las derivaciones de este artículo se basan en la TCT, no se consideran las propiedades distintivas de esta teoría, los modelos de la teoría de medición de Rasch (RMT) y los modelos de la teoría de respuesta al ítem (IRT). En lugar de ello, el modelo dicotómico de Rasch de la RMT se emplea solo para simular las respuestas, dejando claro que sus propiedades lo hacen relevante para este fin. Usando la nomenclatura de la TCT para la puntuación real de una persona, la probabilidad de obtener una respuesta dicotómica $x = 0,1$ en el modelo está dada por

$$\Pr\{X_{ni} = x; \beta_n, \delta_i\} = \frac{\exp(\tau_n - \delta_i)}{1 + \exp(\tau_n - \delta_i)}, \quad (33)$$

donde τ_n es la puntuación real de una persona n, y δ_i es la dificultad del ítem i. En la simulación de las propiedades de las subescalas, la estructura de las subescalas de la ecuación (13) fue usada para calificar τ_n. Así, más que usar la ecuación (19) de la TCT para generar una puntuación observada $X_{ni} = x$ sobre la base de una estructura paramétrica aditiva y un componente aleatorio de la ecuación (13), se emplea el componente aleatorio probabilístico no lineal de la ecuación (33) en la misma estructura paramétrica para generar las puntuaciones observadas. De este modo, dentro de cada subescala, se satisface el modelo dicotómico de Rasch y, entre las subescalas, la correlación latente global es la perteneciente al ejemplo real.

La Prueba de Aptitud Académica de Australia

La Prueba de Aptitud Académica de Australia (Australian Scholastic Aptitude Test, ASAT) es un test de opción múltiple con 100 ítems construido para cubrir cuatro áreas de logro académico: Matemáticas, Ciencias, Humanidades y Ciencias Sociales. Generalmente se usa a nivel del grado 12 para evaluar a los alumnos que postulan para ingresar a las universidades en Australia. La muestra completa de las respuestas de la ASAT analizadas en este artículo fue usada para la equiparación a una persona común de todas las asignaturas del año 12 (por ejemplo: Literatura Inglesa, Historia, Matemáticas, Física) que estudian los alumnos para prepararse para la selección universitaria. Como incentivo para que los estudiantes respondan la prueba con seriedad, un pequeño porcentaje de la puntuación total de cada estudiante en la ASAT se añadió a su puntuación de ingreso a la universidad. Con este propósito, y con el fin de realizar escalamiento, se utilizó la puntuación total en la ASAT.

El número de estudiantes cuya información quedó disponible corresponde a una muestra aleatoria de 1.000 sujetos. Sin embargo, 13 de los estudiantes no entregaron respuestas completas y por lo tanto quedaron fuera del análisis. La muestra restante constó de 490 alumnas y 497 alumnos. Las puntuaciones máximas en las cuatro escalas fueron desde 23 hasta 27 (Matemáticas, 27; Ciencias Naturales, 23; Ciencias Sociales, 24; Humanidades, 26). Por consiguiente, para la estimación de c se usó la ecuación (27) en lugar de la (26).

Aunque las áreas sustantivas de Matemáticas, Ciencias Naturales, Ciencias Sociales y Humanidades son claramente diferentes, todos los estudiantes que respondieron esta prueba se estaban preparando para estudiar en universidades ubicadas en Australia Occidental o en otras partes del país. Por lo tanto, estaban estudiando cada una de estas disciplinas. Los ítems seleccionados no se enfocaban en un currículo específico de grado 12, aunque se esperaba que los estudiantes tuvieran una preparación correspondiente a los estudios de este nivel. Por lo tanto, una vez más, aparece como razonable la hipótesis de que los perfiles de los estudiantes en las cuatro subescalas pueden englobarse en un número único, lo que es evidencia de unidimensionalidad. Sin embargo, también existe una clara estructura de subescala que evalúa diferentes aspectos de la variable de la *aptitud académica*. La Tabla 2 muestra los resultados del análisis del ASAT de acuerdo a las fórmulas presentadas más arriba y de la media y la desviación estándar del valor de c en el estudio de simulación, el valor estadístico sobre el cual se basan todas las demás estimaciones.

Tabla 2
Tres cálculos de α, correlación de las subescalas en ASAT y un resumen de 10 réplicas.

α_c	$\alpha_{S,}$	α_A	ρ
0.924	0.819	0.886	0.658
c	Valor observado		0.722
10 Réplicas		Media	0.719
		DesvEst	0.027

En primer lugar, la Tabla 2 muestra que $\alpha_c > \alpha_S$, en segundo lugar, que la estimación de c es relativamente grande (0.722), y en tercer lugar, sobre la base de las desviaciones estándar de los valores de c para las 10 réplicas, que el valor observado de 0.772 para c es significativamente mayor que 0. Desde la perspectiva de la estabilidad de las estimaciones, es claro que el valor de generación de 0.722 se halla perfectamente dentro de los límites de confianza de 95% indicados por la desviación estándar de los valores estimados de c de acuerdo a las 10 réplicas.

La Tabla 2 también muestra que, debido a un valor relativamente grande para c (0.722), la correlación global latente entre diferentes subescalas es relativamente baja: 0.658. Sin embargo, α_A (la proporción de la varianza común de la escala en relativa a las varianzas común y única, es decir, la proporción de la varianza de no error), es relativamente alta: 0.9 (0.886). Esto indica que las subescalas estaban, en general, correlacionadas en un grado suficientemente alto de modo que, junto con las cuatro subescalas, el mayor componente de la varianza era la varianza común verdadera. Recuérdese que, según la ecuación (22), cada subescala adicional suma cuadráticamente a la varianza común pero solo linealmente a la varianza única.

Esta alta proporción de varianza común verdadera de la escala sugiere que, para los propósitos que se buscan alcanzar con ASAT, emplear una puntuación única era en general justificable. Para consolidar la interpretación, relacioné los resultados de la Tabla 2 con algunas de las derivaciones. En particular, en lugar de considerar solo proporciones de varianza, es posible estimar los componentes de la varianza común, de la varianza única y de las varianzas de error en términos de la varianza de las puntuaciones totales observadas. Con esta intención se emplearon las relaciones presentadas en la Tabla 1.

En primer lugar, del valor de α_S y el valor de la varianza total σ_y^2 obtenidos mediante el análisis, puede hacerse fácilmente una estimación de la varianza verdadera de la escala, σ_t^2. Específicamente, $\sigma_t^2 = \alpha_S \sigma_y^2$. En segundo lugar, sobre la base del valor de α_c y su expresión, y el valor disponible de σ_t^2 con valores conocidos de S, K, puede calcularse la estimación de la varianza única, σ_u^2. Finalmente, dados estos valores, sobre la base de la ecuación (23), puede calcularse la varianza de error. La Tabla 3 muestra los resultados de estos cálculos. Nuevamente queda claro que el componente dominante de la varianza es la varianza común (212.523) y que la varianza única es relativamente pequeña (28.102) y no mucho mayor que la relativamente pequeña varianza de error (18.873).

Tabla 3
Componentes de la varianza de ASAT en la escala de las puntuaciones observadas

σ_y^2	σ_t^2	σ_u^2	σ_e^2
259.528	212.523	28.102	18.873

Considerando la estructura a priori de las subescalas y el conocimiento de que, en presencia de escalas, el valor resultante de α puede inflarse, el alto valor de α_c (0.924) por sí solo no habría sido suficiente evidencia para concluir que la puntuación total es un resumen razonable de los perfiles de la mayoría de los estudiantes. Sin embargo, en un contexto en donde los estudiantes están preparados para responder todos los tipos de preguntas, ASAT funcionó de manera relativamente unidimensional, a pesar de que las subescalas estaban compuestas de ítems sustantivos claramente muy diferentes.

Sin embargo, a nivel individual, habría una minoría de estudiantes para quienes la puntuación total no resume el perfil de cuatro puntuaciones. Los perfiles que no pueden resumirse mediante una puntuación única y que contribuyen a una correlación latente menor que 1, es decir, que contribuyen en algún grado a la multidimensionalidad en los datos como un todo, podrían identificarse, considerándose esas personas como casos especiales. Por ejemplo, la persona con el perfil más heterogéneo fue una estudiante cuyas puntuaciones en matemáticas, ciencias, ciencias sociales y humanidades fueron 26, 4, 7 y 7 respectivamente. Esta estudiante claramente tuvo un desempeño excelente en matemáticas y relativamente débil en las otras subescalas. Es probable que haya sido una hablante nativa de una lengua distinta al inglés. En el momento de la administración de la prueba ASAT se tuvo cuidado de remover las puntuaciones de los hablantes de inglés como segunda lengua cuando se utilizó la prueba con el propósito de equiparar puntuaciones en otras áreas disciplinares.

Discusión

El presente artículo deja de manifiesto que, a pesar de las muchas advertencias sobre la interpretación del coeficiente α , este seguirá siendo usado por los investigadores en escalas en las cuales resumen a la persona que responde la prueba con una puntuación única sumada en los ítems. Esto ocurre en parte porque α es fácil de calcular sobre la base de una sola administración de una prueba y es un índice de confiabilidad dentro de la TCT si los datos cumplen ciertos supuestos relativamente fuertes.

Este artículo también indica que muchas escalas se construyen de modo de tener subescalas que evalúen diferentes aspectos de la variable común, y que por lo tanto las subescalas tendrán varianzas únicas. Entonces, en lugar de enfocarse en las circunstancias en las cuales el índice α se inflará o conducirá a error de otras formas ya bien descritas en la literatura, el artículo califica su aplicación y muestra que, sobre la base de dos cálculos de α , uno a nivel de los ítems originales y otro a nivel de las subescalas, pueden estimarse varios componentes y proporciones de la varianza. Se sugiere que al reportar estos componentes e interpretarlos en el contexto de una escala específica dentro de un contexto específico, es más probable que se obtengan interpretaciones exhaustivas y precisas, de preferencia a emplear un índice único para todas las circunstancias o abandonar del todo el uso de α .

Este artículo muestra que al trabajar con algunos principios subyacentes, una estructura bifactorial, y algunos supuestos específicos de aceptación generalizada, el cálculo de α a nivel de los ítems entrega la proporción de las varianzas verdaderas común y única relativas a la varianza total, y si se calcula a nivel de las subescalas, entrega la proporción de la varianza común verdadera respecto de la varianza total. Dos índices que califican la interpretación de estos dos cálculos de α se desprenden inmediatamente de este trabajo: el primero, una correlación latente entre las subescalas; y el segundo, la proporción de la varianza común verdadera de la escala relativa a la suma de esta varianza y la varianza única verdadera de las subescalas.

El enfoque de usar los índices se demuestra mediante un ejemplo empírico, que cuenta con una clara estructura de subescalas, en la que existe una predisposición a usar una puntuación única para resumir el perfil de cada estudiante en la escala, y donde por lo tanto se presume implícitamente la presencia de una variable unidimensional. Mientras mayor sea el valor de cada uno de los tres índices de α – los cuales reflejan proporciones de diferentes componentes de la varianza común, única y de error – y de la correlación latente entre las subescalas, menor será el número de perfiles para los cuales el uso de la puntuación total es inconveniente. Aquellos que no pueden ser englobados en una puntuación total podrían ser los perfiles más importantes para considerar desde una perspectiva sustantiva o clínica. Un pequeño estudio de simulación con una estructura de subescala que hizo el paralelismo con datos empíricos mostró la estabilidad de las estimaciones de los índices desarrollados en el artículo.

Por supuesto, hay limitaciones para la interpretación de los índices que se mencionan en este artículo y que necesitan ser destacadas. En primer lugar, debido a la presunción de continuidad de las respuestas en su derivación, lo que se aproxima por respuestas típicamente discretas, la distribución de las personas no debería tener un sesgo artificial que influyera en producir grandes cantidades de sujetos con puntuaciones mínimas o máximas. En segundo lugar, aunque los cálculos puedan llevarse a cabo con un número variable de ítems por subescala, las cantidades de ítems (o las puntuaciones máximas cuando los ítems dentro de las subescalas se suman) debieran ser relativamente similares; de no ser así, las subescalas con números mayores de ítems dominan los índices y podrían llevar a interpretaciones erróneas. En el ejemplo, las puntuaciones fueron desde 23 hasta 27 en cada subescala, y la simulación que tuvo exactamente el mismo número de ítems en las subescalas mostró que con esta variación de ítems, se tuvieron predominantemente estimaciones estables. En tercer lugar, relacionada con las limitaciones presentadas previamente, está la presunción de que las varianzas son relativamente homogéneas entre las subescalas. Por supuesto, esta presunción puede comprobarse empíricamente. En cuarto lugar, el supuesto de que las correlaciones entre todos los pares de subescalas son homogéneas también es fuerte. Esto también puede comprobarse empíricamente. En quinto lugar, idealmente la estructura de las subescalas se conoce con antelación. En algunos casos, es posible derivar una estructura de subescala sobre la base de los datos mismos empleando varias técnicas, como el análisis factorial. Sin embargo, definir subescalas basadas en dichos enfoques, pueden inducir un riesgo de depender de la aleatoriedad de los datos.

El uso de estos índices y sus interpretaciones en cualquier conjunto de datos con respecto de los cuales se consideren relevantes no impide que se realicen otros cálculos e investigaciones. Por ejemplo, pueden emplearse los propuestos por McDonald (1978), en los cuales diferentes subescalas pueden tener pesos distintos, considerando además interpretaciones de confiabilidad. Otro enfoque, basado en modelos de ecuaciones estructurales, que considera diferentes pesos para las subescalas, fue propuesto por Raykov y Shrout (2002); y, un tercer enfoque, basado en una estructura bifactorial como en el presente artículo (Reise et al., 2007; Reise et al., 2010), aplica los principios de la teoría de respuesta al ítem. En muchas de estas aplicaciones, el concepto de confiabilidad se mantiene. En cambio, en la teoría de la generalizabilidad (Brennan, 1997) el foco está en los componentes de la varianza. Las implicaciones del presente artículo son compatibles con la comprensión de una escala con subescalas desde el punto de vista de los componentes de la varianza.

Finalmente, dada su historia, su ubicuidad y la facilidad con la que puede calcularse, es probable que α siga siendo usado e interpretado, tomando en cuenta, más o menos seriamente, las advertencias presentes en la literatura. Cada uno de los índices descritos en este artículo, junto con otras fórmulas relacionadas y que resumen los componentes de la varianza de las subescalas de una escala, puede calcularse empleando otros procedimientos, incluyendo la teoría de la generalizabilidad, pero se sugiere que es probable que se utilicen cuando se calculen sobre la base de solo dos valores de α. Como producto adicional de este trabajo, puede afirmarse que, a pesar de las limitaciones, los errores de interpretación respecto de α podrían minimizarse y, en particular, contrarrestar la impresión de que este coeficiente es un índice de unidimensionalidad.

El artículo original fue recibido el 20 de julio de 2015
El artículo revisado fue recibido el 20 de agosto de 2015
El artículo fue aceptado el 24 de agosto de 2015

Referencias

Andersen, E. B. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42,* 357-74.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-574.

Brennan, R. L. (1997). A perspective on the history of generalizability theory. *Educational Measurement: Issues and Practice, 16,* 14-20,

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189-225.

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, 16,* 297-334.

Gibbons, R. D., & Hedecker, R. D. (1992). Full information bi-factor analysis. *Psychometrika, 57,* 423-436.

Green, S. B., Lissitz, R. W., & Muliak, S. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-833.

Gulliksen, H. (1950). *Theory of mental tests.* Nueva York: Wiley.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255-282.

Holzinger, K. J., & Swineford, F. (1937). The Bi-factor method. *Psychometrika, 2,* 41-54.

Komaroff, E. (1997). Effect of simultaneous violations of essential tau equivalence and uncorrelated errors on coefficient alpha. *Applied Psychological Measurement, 21*, 337-348.

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

Marais, I., & Andrich, D. (2008). Formalising dimension and response violations of local independence in the unidimensional Rasch model*.* *Journal of Applied Measurement, 9,* 200-215.

McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability". *Educational and Psychological Measurement, 38,* 75-79.

Mehrens, W. A., & Lehman, I. J. (1991). *Measurement and evaluation in education and psychology.* (4th ed.). Nueva York: Harcourt Brace.

Rae, G. (2006). Correcting coefficient alpha for correlated errors: Is α_k a lower bound for reliability? *Applied Psychological Measurement, 30*, 56-59.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. En J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. IV* (pp. 321-334). Berkeley CA: University of California Press.

Raykov, T. (1998). Coefficient Alpha and the composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22,* 375-385.

Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 195-212.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*, 544-559. doi: 10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19-31.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of coefficient alpha. *Psychometrika, 74,* 107-120. doi: 10.1007/s11336-008-9101-0

Smith, E. (2005). Effect of Item redundancy on Rasch Item and person estimates. *Journal of Applied Measurement, 6,* 147-163.

Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of coefficient Alpha. *Psychometrika*, *65,* 271-280.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement.* Chicago: MESA Press.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the medical college admissions test. *Journal of Educational Measurement, 39*, 291-309.

Zhang, J., & Stout, W. (1999). Theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213-249.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123-133. doi: 10.1007/s11336-003-0974-7

Apéndice A

Cálculo de una estimación de α

El apéndice resume una derivación de α que comienza con las varianzas a nivel de los ítems, la cual es usada luego para formalizar la construcción de α cuando se toma en consideración la estructura de las subescalas. Los supuestos se entregan en el cuerpo del artículo.

Sobre la base de la ecuación (1) en el texto

$$x_{ni} = \tau_n + \varepsilon_{ni}, \tag{A1}$$

$$V[x_i] = \sigma_\tau^2 + \sigma_\varepsilon^2, \tag{A2}$$

Sobre la base de las ecuaciones (5), (6) y (16) en el texto

$$y_n = \sum_{i=1}^{I} x_{ni} = \sum_{i=1}^{I} (\tau_n + \varepsilon_{ni}) = I\tau_n + \sum_{i=1}^{I} \varepsilon_{ni}, \tag{A3}$$

$$\begin{aligned} V[y] = V[\sum_{i=1}^{I} x_{ni}] &= I^2\sigma_\tau^2 + I\sigma_\varepsilon^2 \\ &= \sigma_t^2 + \sigma_e^2. \end{aligned} \tag{A4}$$

Por lo tanto,

$$\begin{aligned} \frac{V[\sum_{i=1}^{I} x_i] - \sum_{i=1}^{I} V[x_i]}{V[y]} &= \frac{I^2\sigma_\tau^2 + I\sigma_\varepsilon^2 - (I\sigma_\tau^2 + I\sigma_\varepsilon^2)}{I^2\sigma_\tau^2 + I\sigma_\varepsilon^2} \\ &= \frac{\sigma_t^2 - \sigma_t^2 / I}{\sigma_t^2 + \sigma_e^2} \\ &= \frac{I\sigma_t^2 - \sigma_t^2}{I(\sigma_t^2 + \sigma_e^2)} \\ &= \frac{(I-1)\sigma_t^2}{I(\sigma_t^2 + \sigma_e^2)}, \end{aligned} \tag{A5}$$

lo cual lleva a

$$\frac{I}{I-1}\frac{V[\sum_{i=1}^{I} x_{ni}] - \sum_{i=1}^{I} V[x_i]}{V[y]} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_e^2}, \tag{A6}$$

donde el lado izquierdo de la ecuación (A1.7) es el cálculo tradicional del coeficiente α .

El numerador de $N[\alpha]$ de α

Puesto que el denominador en todas las expresiones es simplemente la varianza de las puntuaciones totales, y para simplificar las relaciones entre diferentes cálculos de α en presencia de una estructura de subescalas, me enfoco en el numerador de la expresión para α . Por lo tanto, sobre la base de A1.7, es posible escribir

$$\alpha = N[\alpha]/V[y] \tag{A7}$$

donde $N[\alpha] = \sigma_t^2$. (A8)

Apéndice B

La estructura de las subescalas y las relaciones entre las varianzas

Los supuestos adicionales al considerar la estructura de las subescalas se presentan en el cuerpo del artículo. Aunque $\sigma^2_{\vartheta s} = \sigma^2_{\tau}$; $c_s = c, \forall s$, para efectos de exposición la identidad de la varianza de las subescalas se mantiene hasta el punto de la exposición donde se aplica una suma que requiere la identidad con fines de simplificación.

Varianza de la puntuación total en presencia de una estructura de subescalas

Sobre la base de la ecuación (19) $x_{nis} = \tau_n + c_s\vartheta_{ns} + \varepsilon_{nis}$.

Por lo tanto,

$$V[x_{is}] = \sigma^2_{\tau} + c^2_s\sigma^2_{\vartheta s} + \sigma^2_{\varepsilon}, \tag{B1}$$

y

$$\begin{aligned}\sum_{s=1}^{S}\sum_{i=1}^{K}V[x_{is}] &= \sum_{s=1}^{S}\sum_{i=1}^{K}(\sigma^2_{\tau} + c^2_s\sigma^2_{\vartheta s} + \sigma^2_{\varepsilon}) \\ &= SK\sigma^2_{\tau} + \sum_{s=1}^{S}Kc^2_s\sigma^2_{\vartheta s} + SK\sigma^2_{\varepsilon}. \\ &= SK\sigma^2_{\tau} + \sum_{s=1}^{S}Kc^2\sigma^2_{\vartheta} + SK\sigma^2_{\varepsilon}\end{aligned} \tag{B2}$$

Es decir,

$$\sum_{s=1}^{S}\sum_{i=1}^{K}V[x_{is}] = SK\sigma^2_{\tau} + SKc^2\sigma^2_{\vartheta} + SK\sigma^2_{\varepsilon}. \tag{B3}$$

Varianza de la puntuación de una subescala

$$y_{ns} = \sum_{i=1}^{k}x_{nis} = \sum_{i=1}^{K}(\tau_n + c_s\tau_{ns} + e_{nis}) = K\tau_n + Kc_s\tau_{ns} + \sum_{i=1}^{K}e_{nis}. \tag{B4}$$

Por lo tanto

$$\begin{aligned}V[y_s] &= V[K\tau] + V[Kc_s\vartheta_s] + V[\sum_{i=1}^{K}\varepsilon_{is}] \\ &= K^2\sigma^2_{\tau} + K^2c^2_s\sigma^2_{\vartheta s} + K\sigma^2_{\varepsilon}\end{aligned} \tag{B5}$$

Además, sobre la base de

$$\begin{aligned}y_n = \sum_{s=1}^{S}y_{ns} = \sum_{s=1}^{S}\sum_{i=1}^{K}x_{nis} &= \sum_{s=1}^{S}\sum_{i=1}^{K}(\tau_n + c_s\vartheta_{ns} + \varepsilon_{nis}) \\ &= SK\tau_n + \sum_{s=1}^{S}Kc_s\vartheta_{ns} + \sum_{s=1}^{S}\sum_{i=1}^{K}\varepsilon_{nis},\end{aligned} \tag{B6}$$

$$\begin{aligned} V[y] = V[\sum_{s=1}^{S} y_s] = V\sum_{s=1}^{S}\sum_{i=1}^{K} x_{nis} &= V[SK\tau_n + \sum_{s=1}^{S} Kc_s\vartheta_{ns} + \sum_{s=1}^{S}\sum_{i=1}^{K}\varepsilon_{nis}] \\ &= S^2K^2\sigma_\tau^2 + \sum_{s=1}^{S} K^2c^2\sigma_\vartheta^2 + \sum_{s=1}^{S}\sum_{i=1}^{K}\sigma_\varepsilon^2 \\ &= S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2. \end{aligned} \tag{B7}$$

donde $\sigma_u^2 = SK^2c_s^2\sigma_t^2$ es la varianza única total de todas las subescalas.

La correlación latente entre dos ítems de diferentes subescalas s y u (independiente del error)

Sean K_s, K_w, los números de ítems en las subescalas s y w respectivamente. Entonces, las puntuaciones totales de las subescalas (sin error) son

$$K_s\beta_{ns} = K_s\tau_n + K_sc_s\vartheta_{ns}, \quad K_w\beta_{nw} = K_w\tau_n + K_wc_w\tau_{nw}.$$

Por lo tanto,

$$\begin{aligned} COV[K_s\beta_{ns}, K_w\beta_{nw}] &= COV[K_s\tau_n + K_sc_s\vartheta_{ns},\ K_u\tau_n + K_wc_w\vartheta_{nw}] \\ &= K_sK_w\sigma_\tau^2, \end{aligned} \tag{B8}$$

y la correlación latente ρ_{sw} entre dos subescalas s y w está dada por

$$\begin{aligned} \rho_{SU} &= \frac{K_sK_w\sigma_\tau^2}{\sqrt{K_s^2\sigma_\tau^2 + K_s^2c_s^2\sigma_{\vartheta s}^2}\sqrt{K_w^2\sigma_\tau^2 + K_w^2c_w^2\sigma_{\vartheta w}^2}} \\ &= \frac{K_sK_w\sigma_\tau^2}{K_sK_w\sqrt{\sigma_\tau^2 + c_s^2\sigma_{\vartheta s}^2}\sqrt{\sigma_\tau^2 + c_w^2\sigma_{\vartheta w}^2}} \\ &= \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + c_s^2\sigma_s^2}\sqrt{\sigma_\tau^2 + c_u^2\sigma_u^2}}. \end{aligned} \tag{B9}$$

Con los supuestos $\sigma_\tau^2 = \sigma_{\vartheta s}^2 = \sigma_{\vartheta w}^2;\ c_s = c_w = c,\ \forall s, w$

$$\begin{aligned} \rho_{sw} = \frac{COV[\beta_s, \beta_u]}{\sqrt{V[\beta_s]}\sqrt{V[\beta_w]}} &= \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + c^2\sigma_\tau^2}\sqrt{\sigma_\tau^2 + c^2\sigma_\tau^2}}, \\ &= \frac{1}{1+c^2}. \end{aligned} \tag{B10}$$

Apéndice C

Cuatro cálculos de α para cuatro condiciones diferentes

1. α *: Sin tomar en cuenta una estructura de subescalas cuando* $c_s = 0$

Cuando $c_s = 0$ y se ignora cualquier estructura de subescalas, se está en presencia del caso estándar para el cálculo de α. Sin embargo, para comparar las fórmulas bajo diferentes condiciones, debe tenerse en cuenta de que existen *S* subescalas con *K* ítems cada una. Por lo tanto, $I = SK$ y usando la ecuación (A1.9), el numerador $N[\alpha]$ está dado por

$$N[\alpha] = I^2\sigma_\tau^2 = S^2K^2\sigma_\tau^2 = \sigma_t^2, \tag{C1}$$

y

$$\alpha = \sigma_t^2 / V[y]. \tag{C2}$$

2. α_0 *: Tomando en cuenta una estructura de subescalas cuando* $c = 0$

Para calcular el numerador $N(\alpha_0)$ cuando se toma en consideración la estructura de las subescalas, la suma de los ítems dentro de cada subescala se usa para puntuar *S* subescalas.

Aplicando la estructura de las ecuaciones (A2.5) y (A2.7) donde las subescalas reemplazan ítems, el numerador $N[\alpha_0]$ toma la forma

$$\begin{aligned} N[\alpha_0] &= \frac{S}{S-1}(V[\sum_{s=1}^{S} y_s] - \sum_{s=1}^{S} V[y_s]) \\ &= \frac{S}{S-1}(S^2K^2\sigma_\tau^2 + SK^2c_s^2\sigma_s^2 + SK\sigma_\varepsilon^2 - S(K^2\sigma_\tau^2 + K^2c_s^2\sigma_s^2 + K\sigma_\varepsilon^2)) \\ &= \frac{S}{S-1}(S^2K^2\sigma_\tau^2 - SK^2\sigma_\tau^2) \\ &= \frac{S}{S-1}SK^2\sigma_\tau^2(S-1) \\ &= S^2K^2\sigma_\tau^2 = I^2\sigma_\tau^2 \\ &= \sigma_t^2. \end{aligned} \tag{C3}$$

Por lo tanto,

$$\alpha_0 = \sigma_t^2 / V[y]. \tag{C4}$$

Es relevante apuntar que en este caso el valor de c no tiene ninguna función en el numerador, sea 0 o no. Sin embargo, puesto que es 0, tampoco tiene ninguna función en el denominador, donde, sobre la base de la ecuación (A2.7), el denominador se reduce a $V[y] = \sigma_t^2 + \sigma_e^2$.

3. α_c: Sin tomar en consideración una estructura de subescalas cuando $c_s \neq 0$

Si $c_s \neq 0$ y la estructura de las subescalas no se toma en cuenta, entonces hay *SK* ítems discretos. Al aplicar las ecuaciones (A2.3) y (A2.7),

$$\begin{aligned} N[\alpha_c] &= \frac{SK}{SK-1}(V[\sum_{s}^{S}\sum_{i=1}^{K} x_{is}] - \sum_{s=1}^{S}\sum_{i=1}^{K} V[x_{is}]) \\ &= \frac{SK}{SK-1}(S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2 - (SK\sigma_\tau^2 + SKc^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2)) \\ &= \frac{SK}{SK-1}(S^2K^2\sigma_\tau^2 - SK\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 - SKc^2\sigma_\vartheta^2) \\ &= \frac{SK}{SK-1}(SK(SK-1)\sigma_\tau^2 + SK(K-1)c^2\sigma_\vartheta^2 \\ &= S^2K^2\sigma_\tau^2 + \frac{S^2K^2(K-1)c^2\sigma_\vartheta^2}{SK-1} \\ &= I\sigma_\tau^2 + \frac{(SK^2c^2\sigma_\vartheta^2)S(K-1)}{SK-1} \\ &= \sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1). \end{aligned} \tag{C5}$$

Por lo tanto,

$$\alpha_c = [\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1))/V[y]. \tag{C6}$$

4. α_S: Tomando en cuenta una estructura de subescalas cuando $c_s \neq 0$

Tomar en cuenta la estructura de las subescalas cuando $c_s \neq 0$ de nuevo simplemente implica sumar los ítems dentro de una escala para dar *S* ítems de orden superior. Al aplicar las ecuaciones (A2.5) y (A2.7) a estos *S* ítems se obtiene

$$\begin{aligned} N[\alpha_S] &= \frac{S}{S-1}(V[\sum_{s=1}^{S} y_s] - \sum_{s=1}^{S} V[y_s]) \\ &= \frac{S}{S-1}(S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2 - S(K^2\sigma_\tau^2 + K^2c^2\sigma_\vartheta^2 + K\sigma_\varepsilon^2)) \\ &= \frac{S}{S-1}(S^2K^2\sigma_\tau^2 - SK^2\sigma_\tau^2) \\ &= \frac{S}{S-1}SK^2\sigma_\tau^2(S-1) \\ &= S^2K^2\sigma_\tau^2 = I^2\sigma_\tau^2 \\ &= \sigma_t^2. \end{aligned} \tag{C7}$$

Por lo tanto,

$$\alpha_S = \sigma_t^2 / V[y]. \tag{C8}$$