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

Pontificia Universidad Católica de Chile
Facultad de Educación

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

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

David Andrich

Graduate School of Education, The University of Western Australia

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

Post to:
David Andrich
Graduate School of Education, The University of Western Australia
M428, 35 Stirling Highway, Crawley, Western Australia , 6009
Email: david.andrich@uwa.edu.au
An earlier version of this paper was presented at an International Objective Measurement Workshop, New York, NY, April 2008, and at an International Conference on Outcomes Measurement, Bethesda, Maryland, September, 2010. The research reported in this paper was supported in part by Australian Research Council Linkage grants. Permission to use data from the Australian Scholastic Aptitude Test, which is a test constructed by the Australian Council for Educational Research, was given by the former Curriculum Council of Western Australia, now the School Curriculum and Standards Authority.

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

Resumen

En las mediciones sociales, muchas escalas construidas para medir una sola variable 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 resumida entre subescalas inmediatamente corregida para la atenuación debida al error. El artículo sugiere que, puesto que los cálculos son relativamente simples, pueden usarse para entregar un resumen de las propiedades de una escala con subescalas más amplio de lo que permite una única estadística, como por ejemplo, alguna forma de coeficiente de confiabilidad. Este artículo entrega un ejemplo en el cual un test de aptitud académica, consistente en 100 ítems, 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 la varianza

Many scales in psychology, education, health assessment, and social measurement in general, are constructed to measure a single variable. Such scales are generally said to be *unidimensional.* In all test theories (Lord & Novick, 1968) of unidimensional scales, each person is characterized by a single value and, in Classical Test Theory (CTT), by the total score. Practical guides on test construction include advice on constructing items that are most likely to be statistically independent (Mehrens & Lehman, 1991) in assessing the same variable. The two main reasons that scales are composed of more than one statistically independent item are: one, the greater the number of items, the greater the number of potential total scores and therefore the greater the potential precision; and two, the greater the potential number of *aspects* of the same variable that are assessed, the greater the potential for validity.

Many scales constructed to be unidimensional also have more than one item to assess each of multiple aspects of the variable. The scale then explicitly has *subscales of items*. Although by definition all subscales assess a common variable, each also assesses a variable unique to its aspect. As an artifact, the scale is technically no longer unidimensional. Nevertheless, because together they are considered to assess the common variable more validly than if there were only one item per aspect, the subscales with multiple items are retained.

The presence of subscales lends itself to hypothesising a simple bifactor structure (Holzinger & Swineford, 1937), which has been reconsidered in related contexts (Chen, West, & Sousa, 2006; Gibbons & Hedecker, 1992; Raykov & Shrout, 2002; Reise, Moore, & Haviland, 2010; Reise, Morizot, & Hays, 2007; Zhang & Stout, 1999), and therefore to the analysis of the relative components of variance accounted for by the common and unique aspects. Working from principles of CTT and the bifactor structure in which homogeneous unique variances for subscales and homogeneous covariances among all pairs of subscales is assumed, together with common simplifying assumptions of homogeneous and uncorrelated error variances, this paper shows that from the two calculations of the familiar coefficient α , one at the level of the items and one at the level of the subscales, the true variance common to all subscales, the true variance unique to the subscales, and the error variance can be calculated. As a result, four common indices of interest can be calculated: (a) the summary correlation between pairs of subscales corrected immediately for attenuation due to error; (b) the proportion of the total observed variance which is the sum of the common and unique subscale variances; (c) the proportion of the total variance which is the common variance; and (d) the proportion of the common variance relative to the sum of the common and unique variances.

Under certain circumstances, the coefficient α provides an estimate of the reliability of a scale defined in CTT as the ratio of true score to the total variance, the latter variance being the sum of true and error variance. Many pitfalls in interpreting α as a reliability, and recommendations to overcome these pitfalls, have been suggested in the literature (e.g., 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), since α's elaboration by Cronbach (1951) based on Guttman (1945).

The connotations of the term *reliability* are evaluative and positive —the greater the reliability, the better. It is perhaps because of this connotation, together with the attempt to find and suggest the ideal conceptualizations of reliability (e.g., Zinbarg et al., 2005), that there is such a vast literature on the topic. This paper is not concerned with adding to that kind of literature. Instead, this paper is concerned with exploiting the simple calculations of α to provide the components of variance of a scale with a priori subscales when some simplifying assumptions to the subscale structure are made. Then, given this information, I leave it to the researcher, or to the reader, to make evaluative judgments for a particular scale in the relevant context, which might include further analyses. Thus, this paper is concerned only incidentally with the CTT definition of reliability, which is noted in passing. In addition, this paper recognizes that the focus in generalizability theory is on the decomposition of variance (Brennan, 1997). However, it is suggested that because the calculation of α is seemingly ubiquitous and the proposed calculations are relatively simple, they can be used to readily provide a summary of the components of variance of a scale with subscales.

The paper formalizes and exploits the observation that α calculated at the level of the original items is greater than when it is calculated at the level of the subscales (for example, Marais & Andrich, 2008; Rae, 2006; Smith, 2005; Zenisky, Hambleton, & Sireci, 2002). The rest of the paper is structured as follows. Section 2 summarizes the inferences from coefficient α when calculated at the level of the items and the level of the subscales, Section 3 shows the analyses and interpretation of the components of variance of a real data set, and Section 4 is a summary and discussion. Because the paper concentrates on the ratio of α calculated under two conditions, most derivations are provided for completeness in the Appendices.

Formalization of a subscale structure and coefficient α

In anticipation of the formalization of the subscale, with a bifactor structure, I will first summarize the calculation of α where the assumptions of CTT (Gulliksen, 1950; Lord & Novick, 1968) are met and there is no subscale structure. The observed score of a person on an item is resolved first into its true and error components and is then related to the usual similar resolution of the total score on a scale.

CTT variances in the standard case

Let the observed score of person *n* on item i, $i = 1,2,...,I$, be x_{ni}, $x_{ni} \in \{0,1,2,..m_i\}$ where m_i is the maximum score of item i. Let x_{ni} be resolved according to

$$x_{ni} = \tau_n + \varepsilon_{ni} \tag{1}$$

where τ_n is the value on the variable τ common to the responses of person n to all items and ε_{ni} is the error component of this response to item i. τ_n is referred to in this paper as the *common item true score*. (Although the use of the term "common" is strictly redundant here, it is used in anticipation of distinguishing it from a "unique" score in the context of a subscale structure). Although fixed for person n, for a relevant population of persons, τ and ε_i are taken as random variables. Let ε_i be uncorrelated with the person values, and be distributed homogeneously among items and normally with mean 0, variance $N(0,\sigma_\varepsilon^2)$. Although not strictly necessary, for convenience, let the person distribution also be normally distributed. Then

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

and from Eqs. (1) and (2)

$$V[x_i] = V[\tau] + V[\varepsilon], \tag{3}$$

that is $\sigma_i^2 = \sigma_\tau^2 + \sigma_\varepsilon^2$, (4)

where $\sigma_i^2 = V[x_i]$; $\sigma_\tau^2 = V[\tau]$ and $\sigma_\varepsilon^2 = V[\varepsilon]$, in which, because of homogeneity, there is no need for the subscript i. For convenience of exposition, the operator notation $V[.], COV[.]$ is used in derivations and the notation σ^2 for resultant values.

Let the total score of person n on the scale composed of I items be $y_n = \sum_{i=1}^{I} x_{ni}$ giving

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

Further, let

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

Then, from (5) and (6),

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

where t_n is the CTT true score and e_n is the error on the scale as a whole for person n . t_n is referred to as the *common scale true score* to distinguish it from the common item true score τ_n . From Eq. (7),

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

Eqs. (7) and (8) characterize CTT.

From the assumptions that the errors are homogeneous across items and that $COV[\varepsilon_i, \varepsilon_j] = 0$ for all i, j, from Eq. (6),

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

It will prove convenient to use Eq. (9) in the derivations and interpretations of the paper. One immediate observation from Eq. (9) is that although the variance of the errors increases only linearly as a function of the number of items, the variance of the common scale true scores increases quadratically, ensuring that as the number of items increases, this variance increases relative to the error variance.

The traditional (theoretical) reliability, notated ρ_{yy} here, is defined by the ratio

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

which is clearly the proportion of the common scale true score variance, σ_t^2, relative to the observed total score variance σ_y^2.

The variables $\tau_n, \varepsilon_{ni}, t_n, e_n$ are assumed continuous; therefore, the implication is that x_{ni} and y_n are also continuous. However, because x_{ni} generally takes on only integer values, often simply 0 and 1 with dichotomously scored items, y_n also has discrete integer values and the implied continuity in Eq. (7) is violated. I discuss this violation and one major implication later in the paper but, for the present, I assume that, as is done traditionally, the approximation of discrete for continuous scores is satisfactory.

A review of the calculation of coefficient α

For completeness, the derivation of the equation for calculating α , shown in full in Appendix A, begins with the resolution of the response to each item according to Eq. (1). From Appendix A and Cronbach (1951),

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

Eq. (11) shows that when the assumptions of CTT are fully met, then α provides a calculation of the reliability of CTT; otherwise, it is a lower bound (McDonald, 1978). Eq. (11) also shows how the value of $\sigma_e^2 = \sigma_\varepsilon^2 / I$ decreases, and therefore α increases, as the number of items increases. In this paper, I identify α with the *structure of the calculation* of the left side of Eq. (11), and consider the consequences of such a calculation in the presence of subscales.

I use the notation $V[y]$ for the denominator in the calculations of α but to emphasize the structure in Eq. (11), I use $V[\sum_{i=1}^{I} x_i]$ for $V[y]$ in the numerator.

Formalizing the subscales of a scale

The bifactor structure is formalized at the item level, which requires a qualification of Eq. (1). Accordingly, resolve x_{ni}, giving

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

where τ_n retains the same meaning as in Eq. (1), ϑ_{ni} is the value on the unique aspect of the variable of person n to item i, c_i is a weight for item i and ε_{ni} is again the error component in person *n*'s response to item i. If each item assesses a unique aspect which is uncorrelated with the unique aspect of any other item, that is, there is no subscale structure, then the unique aspect of each item is absorbed into the error, $c_i \vartheta_{ni} + \varepsilon_{ni} \rightarrow \varepsilon_{ni}$, resulting in Eq. (1). However, with a subscale structure, and assuming that $c_i = c_s$ is the same value for all items within each subscale, the unique subscale variance can be separated from the error variance. The result of this paper is that, further assuming that $c_s = c$ for all items between every subscale, c can be estimated from two calculations of α. It is clear that specifying $c_i = c_s = c \; \forall i, s$ is a strong simplifying assumption. However, it is the estimation of c under such a simplifying assumption that permits a rapid overview of approximations of the components of variance among the subscales.

In the first instance, and for purposes of exposition, retain the subscript s in c_s. Then, let

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

$c_s \geq 0$, where τ_n is again the common, item true score for person *n* among subscales and is the same variable as in Eq. (1), and ϑ_{ns} is the unique item score for all items in subscale s. For convenience, and because it has no error term, β_{ns} is referred to as a *unique subscale true score*. According to the bifactor structure $COV[\vartheta_s,\tau]=COV[\vartheta_s,\vartheta_w]=0$ for all subscales s,w, giving

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

Again, although not strictly necessary, let ϑ_s be normally distributed in the same population, $\tau_s \sim N(\mu_s,\sigma_{\vartheta s}^2)$. Further, without loss of generality and for purposes of identification, let

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

for all items within all subscales *s*. Eq. (15) conveniently makes the value c_s^2 the proportion of the unique subscale variance $\sigma_{\vartheta s}^2$ relative to the common item true score variance σ_τ^2: $c_s^2 = (\sigma_{\beta s}^2 - \sigma_\tau^2)/\sigma_{\vartheta s}^2 = (\sigma_{\beta s}^2 - \sigma_\tau^2)/\sigma_\tau^2$. In addition, Eq. (15) implies that the variances of items within and among subscales are homogeneous.

Clearly, if $c_S \neq 0$, the case used in this paper, then taking all items of a scale together, unidimensionality is violated. In the derivations, the terms $V[\tau]$ and $V[\vartheta_s]$ are maintained for clarity of exposition, but to simplify expressions, the numerical equivalent, $\sigma_\tau^2 = \sigma_{\vartheta s}^2$, of Eq. (15) is applied in final expressions of derivations. For the same reason, the subscript s in c_s is retained and then $c_s = c$ is applied in final expressions. (Incidentally, $c_S < 0$ can be taken, but because its effects appear as the square c_S^2, $c_s \geq 0$ is imposed for convenience, implying that subscales are not correlated negatively).

Appendix B shows the calculations of the variance components which arise from Eqs. (12) and (13). For simplicity of exposition, assume first that the number of items K is the same for all S subscales; then, $I = SK$ is the total number of items. I show that this assumption can be relaxed.

If $I = SK$, then from Eq. (9),

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

The correlation between items from different subscales

Appendix B shows that ρ_{sw}, the correlation between the subscale true scores of any pair of subscales s and w is

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

Applying $\sigma_{\tau}^{2} = \sigma_{\vartheta s}^{2} = \sigma_{\vartheta w}^{2}; c_s = c_w = c, \forall s, w$, gives

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

Clearly, the larger the value of c, the smaller the correlation between two subscales from different subscales; if $c = 0$, the correlation is 1, unidimensionality prevails, and there is no effective subscale structure. In the case of just two subscales, ρ_{sw} is identical, conceptually and in value, to the correlation between two subscales corrected for attenuation because of error, that is, $\rho_{sw} = 1/(1+c^2) = r_{sw} / \sqrt{r_{ss}}\sqrt{r_{ww}}$ where r_{sw} is an observed correlation between the subscales and r_{ss}, r_{ww} are estimates of reliability of each subscale. Given a single value of c across all pairs of subscales makes ρ_{sw} a generalization of the correlation between two scales, corrected for error, and is a relevant result in and of itself for the purposes of this paper. In addition, in this paper, and to contrast it with an observed or manifest correlation, ρ_{sw} is referred to as the *latent* correlation among subscales.

Accounting for subscales in calculating α

To account for the subscale structure, each subscale $s, s = 1,2,\ldots,S$ takes the role of an item whose score is the sum of the scores of the original items in that subscale. All the assumptions of CTT listed above are maintained at the subscale level. Again, the formalization begins with the resolution of an item score.

Thus, let the observed score x_{nis} of person *n* on item *i* of subscale *s* be resolved according to

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

where ε_{nis} is the error component of item *i* of subscale *s* for person and $COV[\tau_s, \varepsilon_{is}] = 0$. It is assumed that $\varepsilon_{is} \sim N(0, \sigma_{\varepsilon}^2)$ is homogeneous both among items within a subscale and among items from different subscales so that σ_{ε}^2 is subscripted by neither s nor i. Even though it has the same distribution for all items, for purposes of exposition, I retain the subscripts i and s in ε_{is}.

With *K* items per subscale, let person n's observed score on subscale S be

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

Taking the sum of scores for each person within each subscale, and treating the resultant subscale scores as the units of analysis, is the mechanism for taking account of the subscale structure. Then, the total score of person *n* on the scale is

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

Therefore, using the results in Eqs. (A2.4) – (A2.7) of Appendix B gives

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

where

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

is the sum of the unique variances across all subscales. For the rest of the paper, σ_u^2 and σ_t^2 will be referred to respectively as the unique and common variances, with it being understood that these are true variances in the sense that they are not error variances. Note from Eq. (22) that although each additional subscale adds unique variance $\sigma_u^2 = SK^2c^2\sigma_\vartheta^2$ only linearly as a function of S, each additional subscale adds to the common variance $\sigma_t^2 = S^2K^2\sigma_\tau^2$ quadratically.

Accounting for a subscale structure when $c = 0$ and $c \neq 0$ gives different values for α. The equations are derived in Appendix C and are summarized in Table 1, which gives the numerators for four calculations of α. Their identical denominators are shown in the last line of Table 1.

The notation for each calculation of α is also shown in Table 1. The first column in the first row, $c = 0, \rho_{sw} = 1, \forall s, w$ (unidimensionality is satisfied) takes no account of subscales, is the standard case, and is notated simply as α. The second column of the first row is again $c = 0$, which may be an empirical value, but the formula does take account of a subscale structure and is notated α_0. The third column shows the relationship between α and α_0. In the first column of the second row, $c > 0$, but the formula takes no account of subscale structure and is notated α_c. In the second column of the second row, $c > 0, \forall s$, the formula takes account of the subscale structure and is notated α_S. The third column shows the relationship between α_c and α_S.

Table 1
Conditions under which α is calculated and its values

	Not taking account of the subscale structure	Taking account of the subscale structure	Effect on α
Standard case: $c_S = 0$, $\rho_{sw} = 1$	$\alpha = \frac{\sigma_t^2}{\sigma_y^2}$	$\alpha_0 = \frac{\sigma_t^2}{\sigma_y^2}$	$\alpha = \alpha_0$.
$c_S > 0$, $\rho_{sw} < 1$	$\alpha_c = \frac{\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1)}{\sigma_y^2}$	$\alpha_S = \frac{\sigma_t^2}{\sigma_y^2}$	$\alpha_c > \alpha_S$
$\sigma_y^2 = \sigma_t^2 + \sigma_u^2 + \sigma_e^2$			

The first row in Table 1 shows that if $c = 0$ (there is a subscale structure but with *no* unique subscale values) then α has the same value whether or not a subscale structure is taken into account in its calculation. In the data, there may be a slight difference in values from the two calculations due to different capitalizations of random chance relationships between items in some subscales, but it would be statistically 0. Thus, the fact that α calculated in the two different ways produces effectively the same value is ready evidence that, despite a structure in which subscales potentially assess different aspects, the statistical and empirical unidimensionality of the whole scale prevails.

The second row of Table 1 shows that if $c > 0$ (unique subscale aspects), then $\alpha_c > \alpha_S$. This inequality explains the well-known observation mentioned earlier, which is that when there is a subscale structure, α calculated at the item level is greater than when scores within subscales are summed and α is calculated at the subscale level. The greater value of α_c arises from the addition of a factor of σ_u^2, the unique subscale variance, in the numerator. In the case when the subscale structure is accounted for, this unique variance of the subscales is effectively summed with the error variance leaving only an expression involving the true item variance σ_t^2 in the numerator. Because the expressions in the top row do not involve c $(c = 0)$ in either the numerator or the denominator, the values $\alpha = \alpha_0 = \sigma_t^2 / \sigma_y^2$ are identical. However, although the expression in the second cell of the bottom row takes the same form $\alpha_S = \sigma_t^2 / \sigma_y^2$, because the denominator in the bottom row does involve c, $\alpha = \alpha_0 \neq \alpha_S$.

The result $\alpha_c > \alpha_S$, $c > 0$ also makes clear why the standard calculation of α (α_c) does not indicate the degree of unidimensionality of a scale, which is one of the pitfalls in interpreting α. If the calculated values show $\alpha_c > \alpha_S$, then this inequality confirms unique subscale variance. It is now possible to exploit the difference between α_c and α_S.

Recovering c and ρ_{sw}

Rearranging α_c / α_S provides an estimate of c. Applying Eq. (23) gives

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

giving

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

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

From c^2, $\rho_{sw}, \forall s, w$, is obtained directly from Eq. (18).

As evident from Table 1, if $c > 0$, then $\alpha_c > \alpha_S$. Therefore, if $S > 1$, and because $SK - 1 > K - 1$, Eq. (26) implies that if $\alpha_c > \alpha_S$, then $c^2 > 0$, as required. Clearly, if $S = 1$, then α_S cannot be calculated, but of course, in that case there is no need to attempt to calculate it. As indicated earlier, without loss of generality, take $c = +\sqrt{c^2}$.

The requirement that each subscale have the same number of items can be relaxed. Let k_s be the number of items in subscale s. Then it can be shown that

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

Note that if $k_s \neq K$ for some s, $s = 1,2,\ldots,S$, then even if $c = 0$, $\alpha_0 < \alpha$. Thus, with unidimensionality, when the number of items per subscale is different, the value of α is reduced when items are formed into subscales and analyzed as higher order items. However, because with a subscale structure, α_S is reduced even further, Eq. (27) remains correct. Very different numbers of items or score points per subscale will make the scales with substantially more items or score points dominant in the calculations. Therefore, although it is not necessary to have exactly the same number of items or score points per subscale, for c^2 to be interpreted confidently with the assumption of homogeneity of the variances of the subscales, the number of items or scores points per subscale needs to be very similar and reasonably large; for example, 25 or so dichotomous items per subscale or the equivalent number of score points (e.g., 12 polytomous items with a maximum score of 2). In the illustrative example, the number of dichotomous items per subscale ranges from 23 to 27 and the formulae appear to work accurately.

I now elaborate the interpretation of components of variance given the value of c^2.

Simplifying α_c

From α_c in Table 1 and expanding $V[y]$,

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

Taking the limit of $(K-1)/(SK-1)$ with increasing number of items K, gives

$$\begin{aligned} \lim_{K \to \infty} [K-1]/[SK-1] &= \lim_{K \to \infty} [(1-1/K)/(S-1/K)], \\ &= 1/S, \end{aligned} \tag{29}$$

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

Substituting $1/S$ for $[K-1]/[SK-1]$ in Eq. (28) gives

$$\alpha_c < \frac{\sigma_t^2 + \sigma_u^2}{\sigma_t^2 + \sigma_u^2 + \sigma_e^2}. \tag{30}$$

Eq. (30) shows that α_c is the lower bound for the proportion of the sum of the common and unique variances relative to the sum of the common, unique, and error variances (that is, relative to the total variance). For all practical purposes, it is this proportion. For example, when $K = 15,\ S = 2,\ (1-1/K)/(S-1/K) = 0.483 \approx 0.5 = 1/S$. Thus, $\alpha_c \cong (\sigma_t^2 + \sigma_u^2)/(\sigma_t^2 + \sigma_u^2 + \sigma_e^2)$ can be written.

An interpretation of ratio α_S / α_c

Let the ratio α_S / α_c be denoted by α_A. Then,

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

$$= \frac{\sigma_t^2}{\sigma_t^2 + \sigma_u^2 S(K-1)/(SK-1)}. \quad (31)$$

.

Invoking the limit of Eq. (29) gives

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

For practical purposes and with a reasonably large number of items per subscale, $\alpha_A \cong \sigma_t^2 / (\sigma_t^2 + \sigma_u^2)$. The numerator of α_A is clearly the common variance and the denominator is the sum of the common and unique variances, making α_A effectively the upper bound of the proportion of common variance relative to the common and unique variance, that is, the proportion of the non-error variance.

Describing a scale with four indices

As already indicated, there are many caveats in the literature for interpreting any calculation of α as a reliability index. These caveats seem to arise from the use of one value of α, which might be misleading in different ways in various situations. It is suggested, therefore, that one of the complications in reporting and interpreting α in general, and in particular in the presence of subscales, is to expect that a single value should be sufficient to summarize the properties, and especially some kind of reliability, of a scale. This degree of simplification is not possible. Therefore, to provide a comprehensive summary of the properties of a scale from the perspective of components of variance, it is suggested that all four indices derived above, both calculations of α, the proportion of the non-error variance that is common variance, and the latent correlation among subscales should all be reported.

First, α_S, in which the unique subscale variance is absorbed into the error variance, is most analogous to α when assumptions of CTT are met, and can be reported and interpreted usefully as the proportion of the total variance that is common, true variance. Second, α_A, which indicates the proportion of common variance relative to the sum of the common and unique variances (and independent of error variance), can be interpreted usefully in that it indicates the degree of multidimensionality of the scale. It may also be interpreted as a value of α corrected for attenuation because of random errors in the responses to the items. Third, α_c, the value inflated by the unique variance and the one usually calculated, is also informative in the context of multiple indices as the proportion of common and unique variance. Reporting and interpreting all three values, α_S, α_A, and α_c, together with the value of the summary latent correlation ρ among subscales, is likely to reduce misunderstandings that can result from the interpretation of the value of a single α calculated in the presence of subscales. Such reporting and interpreting is shown in the illustrative example.

Properties of the estimate of c

In the derivations, estimates were not distinguished from parameter values, as it is understood that the calculations of α from data provide estimates. Therefore, the calculation of c from either Eq. (25) or (26), which is a function of the ratio of two calculations of α, is also an estimate. In this paper, there is no opportunity to consider the sampling properties of the estimate of c. However, I provide some simulations with the illustrative example of a real data set.

Before concluding this section, I return to the fact that the derivation of α is based on the assumption of continuous variables even though the responses to items of a scale are generally scored with integers and as a result, the total score is an integer. The key constraint that arises from this feature is that the total scores need to provide sufficient precision that they approximate continuity well enough for the formulae to work. It seems that a maximum score on a scale of at least 25, with a distribution of persons ranging from at least three to a maximum of 22, provides sufficient accuracy to recover the indices derived in this paper.

One important caveat is that the distribution must not be skewed artificially by having many persons with a maximum score and many others with scores close to the maximum, or near a score of 0. In other words, it is important that the items and persons not be so poorly aligned that there is a large floor or ceiling effect, which can occur, for example, if a scale is designed for a normal population and administered to a clinical population, or the other way around. However, this is an important constraint in the general in the application of CTT and the interpretation of reliability coefficients, no matter how they are calculated, because, with floor or ceiling effects, the intercorrelations among items are inflated and so too is any reliability coefficient.

Data analysis

This section illustrates the analysis of a real data set. Although this paper does not consider the statistical properties of the estimate of the variable c in the derivations, to indicate the stability of the estimate with the example, a set of data is simulated from a normal distribution with the means, standard deviations, numbers of items per subscale, the range of item difficulties, and the value of c taking the values obtained from the real data. Then, to elucidate the stability of the estimate of c, ten replications are generated and the mean and standard deviation of the estimates of c are reported.

Simulating discrete responses

To parallel the discrete responses in the real example, simulated responses were generated according to the dichotomous Rasch model with the general class of Rasch models (Andersen, 1977; Andrich, 1978; Rasch, 1961; Wright & Masters, 1982). The key assumptions of CTT and of Rasch models, namely local independence of responses, equal discrimination amongst items, and that the total score provides all information about a person's response profile, are identical. One important difference between CTT and the Rasch models is the specification of item parameters (difficulties in proficiency assessment). In the simulations, these parameters were made equivalent to those from the dichotomous Rasch model analysis of the real data with the same number of items in each of the scales as in the real data. Ten replications, with all parameters the same but with a different random seed, were conducted and the mean and standard deviation of the relevant estimates reported. I stress that the simulations are meant only to be illustrative, and not exhaustive. There was no artificial skew in the real and simulated data.

Because the derivations of this paper are based on CTT, it does not consider the distinctive properties of CTT, the models of Rasch measurement theory (RMT), and models of item response theory (IRT). Instead the dichotomous Rasch model of RMT is employed only to simulate the responses, noting that its properties make it relevant for this purpose. Using the nomenclature of CTT for the person true score, the probability of a dichotomous response $x = 0,1$ in the model is given by

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

where τ_n is the true score of person n, and δ_i is the difficulty of item i. In the simulation of subscale properties, the subscale structure of Eq. (13) was taken to qualify τ_n. Thus, rather than Eq. (19) of CTT being used to generate an observed score $X_{ni} = x$ from an additive parametric structure and a random component of Eq. (13), the non-linear probabilistic random component of Eq. (33) is used on the same parametric structure to generate the observed scores. Thus, within each subscale, the dichotomous Rasch model holds, and between the subscales, the summary latent correlation is that of the real example.

The Australian Scholastic Aptitude Test

The Australian Scholastic Aptitude test (ASAT) is a 100 item multiple choice test constructed to cover four areas of scholastic achievement: Mathematics, Science, Humanities, and Social Sciences. It was generally used at the Year 12 level to assess students who are applying to enter universities in Australia. The full sample of ASAT responses analyzed in this paper were used for the common person equating of all Year 12 subjects (e.g., English literature, history, mathematics, physics) studied by students in preparation for university selection. As an incentive to take the test seriously, a small percentage of each student's total score on the ASAT was added to a student's university entrance score. For this purpose, and for the purpose of scaling, the total score on the ASAT was used.

The number of students for whom data were made available is a random sample of 1,000 students. However, 13 of the students did not have complete responses and were therefore left out of the analysis. The remaining sample consisted of 490 girls and 497 boys. The maximum scores on the four scales ranged from 23 to 27 (Mathematics, 27; Natural Science, 23; Social Science 24; and Humanities, 26). Accordingly, in the estimation of c, Eq. (27), rather than Eq. (26), was used.

Although the substantive areas of Mathematics, Natural Science, Social Science, and Humanities are clearly different, the students sitting for this test were all preparing for study at universities in Western Australia or in other parts of Australia. Therefore, they were studying in each of these disciplines. The items chosen were not focused on a specific Year 12 curriculum, although preparation from Year 12 studies was expected. Thus, once again, the hypothesis that students' profiles on the four subscales could be summarized by a single number, which is evidence of unidimensionality, was reasonable. Nevertheless, there is also a clear subscale structure that assesses different aspects of the variable of *scholastic aptitude*. Table 2 shows the results of the analysis of ASAT according to the above formulae and the mean and standard deviation of the value of c in the simulation study, the statistic on which all other estimates are based.

Table 2
Three calculations of α , the subscales' correlation in the ASAT, and a summary of 10 replications

α_c	$\alpha_{S,}$	α_A	ρ
0.924	0.819	0.886	0.658
c	Observed value		0.722
10 Replications		Mean	0.719
		SD	0.027

Table 2 shows first that $\alpha_c > \alpha_S$, second, that the estimate of c is relatively large (0.722), and third, from the standard deviations of values of c for the 10 replications, that the observed value of 0.772 for c is very much significantly greater than 0. From the perspective of the stability of the estimates, it is clear that the generating value of 0.722 is well within the 95% confidence limits indicated by the standard deviation of the estimated values of c from the 10 replications.

Table 2 also shows that because of a relatively large value for c (0.722), the summary latent correlation between different subscales is a relatively low 0.658. Nevertheless, α_A (the proportion of the scale common variance relative to the common and unique variance, that is, the proportion of the non-error variance), is a relatively high 0.9 (0.886). This indicates that the subscales were, in general, correlated sufficiently highly that, together with the four subscales, the greatest component of variance is the true, common variance. Recall that from Eq. (22), each additional subscale adds quadratically to the common variance but only linearly to the unique variance.

This high proportion of true common variance of the scale suggests that for the purposes to which ASAT was put, using a single score was generally justifiable. To consolidate the interpretation, I related the results of Table 2 to some of the derivations. In particular, rather than considering only proportions of variance, it is possible to estimate the components of common variance, unique variance, and the error variances in terms of the variance of the observed total scores. For this purpose, the relationships in Table 1 were used.

First, from the value of α_S and the value of the total variance σ_y^2 from the analysis, an estimate of the scale true variance σ_t^2 is readily available. Specifically, $\sigma_t^2 = \alpha_S \sigma_y^2$. Second, from the value of α_c and its expression, and now a known value of σ_t^2 with known values of S, K, the estimate of the unique variance σ_u^2 can be calculated. Finally, given these values, from Eq. (23), the error variance can be calculated. Table 3 shows the results of these calculations. It is again clear that the dominant component of variance is the common variance (212.523), and that the unique variance is relatively small (28.102) and not much greater than the relatively small error variance (18.873).

Table 3
Components of variance of ASAT in the scale of the observed scores

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

Considering the a priori structure of the subscales, and the knowledge that in the presence of scales the resultant value of α can be inflated, the high value of α_c (0.924) on its own would not have been sufficient evidence to conclude that the total score is a reasonable summary of the majority of the students' profiles. However, in the context where students are prepared to answer all of the types of questions, the ASAT operated relatively unidimensionally, even though the subscales were composed of clearly very different substantive items.

Nevertheless, at an individual level, there would be a minority for whom the total score does not summarize the profile of four scores. Profiles that cannot be summarized by a single score and that contribute to a latent correlation less than 1, that is, contribute to some multidimensionality in the data as a whole, could be identified and those persons considered as special cases. For example, the person with the most non-homogeneous profile was a girl whose scores in mathematics, science, social science and humanities were respectively 26, 4, 7, and 7. This student has clearly performed excellently in mathematics and relatively poorly in the other subscales. She may have been a native speaker of a language other than English. In the application of ASAT at the time that this test was administered, care was taken to remove the scores of second-language speakers when the test was used for the purpose of equating test scores in other discipline areas.

Discussion

This paper notes that despite the many exhortations cautioning the interpretation of coefficient α, it will continue to be used by practitioners with scales in which they summarize a respondent with a single summed score on the items. This use is in part because α is easy to calculate from one administration of a test and is an index of reliability within CTT if the data meet some relatively strong assumptions.

This paper also points out that many scales are constructed to have subscales that assess different aspects of the common variable and therefore that the subscales will have unique variances. Then, instead of focusing on the circumstances in which the index α will be inflated or misleading in other ways already well covered in the literature, it qualifies its application and shows that from two calculations of α, one at the level of the original items and one at the level of the subscales, various components and proportions of variance can be estimated. It is suggested that reporting these components, and interpreting them in the context of a particular scale in a particular context, is more likely to give comprehensive and accurate interpretations, rather than trying to use a single index for all circumstances or abandoning the use of α altogether.

Working with some underlying principles, a bifactor structure, and some specific and standard assumptions, this paper shows that calculated at the level of the items, α gives the proportion of the true common and true unique variance relative to the total variance, and that calculated at the level of the subscales, it gives the proportion of true common variance relative to the total variance. Two indices that qualify the interpretation of these two calculations of α follow immediately: first, a latent correlation among the subscales, and second, the proportion of true common variance of the scale relative to the sum of this variance and the true unique subscale variance.

This approach to using the indices is demonstrated with an empirical example, which has a clear subscale structure in which there is a predisposition to using a single score to summarize each student's profile on the scale, and therefore implicitly assuming essentially a unidimensional variable. The higher the value for each of the three indices of α, which reflect proportions of different components of common, unique, and error variance, and of the latent correlation among the subscales, the smaller the number of profiles that cannot be summarized by a total score. Those that cannot be so summarized may be the most important profiles to consider from a substantive or clinical perspective. A small simulation study with a subscale structure that paralleled the empirical data showed that the estimates of the indices developed in the paper were stable.

There are of course limitations, broached in the paper, to the interpretation of these indices, which need to be emphasized. First, because of the assumption of continuity of responses in its derivation, which is approximated by typically discrete responses, the distribution of the persons should not have an artificial skew in which large numbers of persons have minimum or maximum scores. Second, although calculations can be carried out with varying numbers of items per subscale, the numbers of items (or the maximum scores when items within subscales are summed) should be relatively similar; otherwise, subscales with greater numbers of items dominate the indices and could lead to misinterpretations. In the example, the scores ranged from 23 to 27 on each subscale, and the simulation that had exactly the same number of items in the subscales showed that with this variation of items, stable estimates prevailed. Third, and related to the above constraints, is the assumption that the variances are relatively homogeneous among the subscales. This assumption can of course be checked empirically. Fourth, the assumption that the correlations among all pairs of subscales are homogeneous is also strong. This too can be checked empirically. Fifth, ideally the subscale structure is known in advance. In some cases it may be possible to derive a subscale structure from the data themselves using various techniques, such as factor analysis. However, forming subscales based on such approaches risks capitalizing on chance.

The use of these indices and their interpretations in any data set for which they are deemed relevant do not preclude other calculations and investigations. For example, those proposed by McDonald (1978), where different subscales may have different weights and with interpretations of reliability, can be carried out. Another approach, which is based on structural equation modeling and considers different weights for subscales, has been proposed by Raykov and Shrout (2002), and a third approach, based on a bifactor structure as in this paper (Reise et al., 2007; Reise et al., 2010), applies the principles of IRT. In many of these applications, the concept of reliability is retained. By contrast, in generalizability theory (Brennan, 1997), the focus is on the components of variance. The implications of the present paper are compatible with providing an understanding of a scale with subscales from the perspective of components of variance.

Finally, given its history, its ubiquity, and the ease with which α can be calculated, it seems it will continue to be used and interpreted, taking into account, more or less seriously, the caveats in the literature. Although each of the indices, and related ones, summarizing the components of variance of subscales of a scale described in this paper can be calculated by other rationales, including generalizability theory, it is suggested that these components are more likely to be used if they can be calculated from just two values of α. Taking into account the limitations, as a byproduct, misinterpretations of α may be minimized and, in particular, the impression that it is an index of unidimensionality may be further countered.

The original article was received on July 20th, 2015
The revised article was received on August 20th, 2015
The article was accepted on August 24th, 2015

References

Andersen, E. B. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42,* 357-374.

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.* New 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.). New 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. In 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

Appendix A

Calculation of an estimate of α

The Appendix summarizes a derivation of α which beings with the item level variances which is then used to formalize the construction of α when account is taken of the subscale structure. The assumptions are provided in the body of the paper.

From Eq. (1) in the text

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

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

From Eqs. (5), (6) and (16) in the text

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

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

Therefore

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

from which

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

where, the left side of Eq. (A1.7) is the traditional calculation of coefficient α.

The numerator of $N[\alpha]$ of α

Because the denominator in all expressions is simply the variance of the total scores, and to simplify relationships among different calculations of α in the presence of a subscale structure, I focus on the numerator of the expression for α. Therefore, from A1.7, it is possible to write

$$\alpha = N[\alpha]/V[y] \quad \text{(A7)}$$

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

Appendix B

The subscale structure and the relationships among the variances

The additional assumptions in accounting for the subscale structure are provided in the body of the paper. Although $\sigma^2_{\vartheta s} = \sigma^2_\tau$; $c_s = c, \forall s$, for purposes of exposition the identity of the subscale variance is retained up to the point in the exposition where a summation is applied which requires the identity for purposes of simplification.

Variance of the total score in the presence of a subscale structure

From Eq. (19) $x_{nis} = \tau_n + c_s \vartheta_{ns} + \varepsilon_{nis}$.

Therefore

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

$$\sum_{s=1}^{S}\sum_{i=1}^{K} V[x_{is}] = \sum_{s=1}^{S}\sum_{i=1}^{K} (\sigma^2_\tau + c_s^2 \sigma^2_{\vartheta s} + \sigma^2_\varepsilon)$$

and

$$= SK\sigma^2_\tau + \sum_{s=1}^{S} Kc_s^2 \sigma^2_{\vartheta s} + SK\sigma^2_\varepsilon. \tag{B2}$$

$$= SK\sigma^2_\tau + \sum_{s=1}^{S} Kc^2 \sigma^2_{\vartheta} + SK\sigma^2_\varepsilon$$

That is,

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

Variance of a subscale score

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

Therefore

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

Further, from

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

$$V[y] = V[\sum_{s=1}^{S} y_s] = V\sum_{s=1}^{S}\sum_{i=1}^{K} x_{nis} = V[SK\tau_n + \sum_{s=1}^{S} Kc_s\vartheta_{ns} + \sum_{s=1}^{S}\sum_{i=1}^{K}\varepsilon_{nis}]$$
$$= S^2K^2\sigma_\tau^2 + \sum_{s=1}^{S} K^2c^2\sigma_\vartheta^2 + \sum_{s=1}^{S}\sum_{i=1}^{K}\sigma_\varepsilon^2 \quad \text{(B7)}$$
$$= S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_\vartheta^2 + SK\sigma_\varepsilon^2.$$

where $\sigma_u^2 = SK^2c_s^2\sigma_t^2$ is the total unique variance of all subscales.

The latent correlation between two items from different subscales s and u (independent of error)

Let K_s, K_w, be the number of items in subscales s and w respectively. Then the total subscale scores (without error) are

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

Therefore,

$$COV[K_s\beta_{ns}, K_w\beta_{nw}] = COV[K_s\tau_n + K_sc_s\vartheta_{ns},\ K_u\tau_n + K_wc_w\vartheta_{nw}]$$
$$= K_sK_w\sigma_\tau^2, \quad \text{(B8)}$$

and the latent correlation ρ_{sw} between two subscales s and w is given by

$$\rho_{SU} = \frac{K_sK_w\sigma_\tau^2}{\sqrt{K_s^2\sigma_\tau^2 + K_s^2c_s^2\sigma_{\vartheta s}^2}\sqrt{K_w^2\sigma_\tau^2 + K_w^2c_w^2\sigma_{\vartheta w}^2}}$$
$$= \frac{K_sK_w\sigma_\tau^2}{K_sK_w\sqrt{\sigma_\tau^2 + c_s^2\sigma_{\vartheta s}^2}\sqrt{\sigma_\tau^2 + c_w^2\sigma_{\vartheta w}^2}} \quad \text{(B9)}$$
$$= \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + c_s^2\sigma_s^2}\sqrt{\sigma_\tau^2 + c_u^2\sigma_u^2}}.$$

With the assumptions $\sigma_\tau^2 = \sigma_{\vartheta s}^2 = \sigma_{\vartheta w}^2;\ c_s = c_w = c,\ \forall s, w$

$$\rho_{sw} = \frac{COV[\beta_s, \beta_u]}{\sqrt{V[\beta_s]}\sqrt{V[\beta_w]}} = \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + c^2\sigma_\tau^2}\sqrt{\sigma_\tau^2 + c^2\sigma_\tau^2}},$$
$$= \frac{1}{1+c^2}. \quad \text{(B10)}$$

Appendix C

Four calculations of α for four different conditions

1. *α : Not taking into account a subscale structure when $c_s = 0$.*

When $c_s = 0$ and any subscale structure is ignored, this is effectively the standard case for the calculation of α . However, to compare the formulae under different conditions, it must be considered that there are *S* subscales with *K* items each. Therefore $I = SK$ and using Eq. (A1.9), the numerator $N[\alpha]$ is given by

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

and

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

2. *α_0 Taking into account a subscale structure when $c = 0$*

To calculate the numerator $N(\alpha_0)$ when the subscale structure is taken into account, the sum of items within each subscale is used to give *S* subscales scores.

Applying the structure of Eqs. (A2.5) and (A2.7) where subscales replace items, the numerator $N[\alpha_0]$ takes the form

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

Therefore,

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

It is relevant to note that in this case, the value of c plays no role in the numerator, whether it is 0 or not. However, because it is 0, it also plays no role in the denominator where, from Eq. (A2.7), the denominator reduces to $V[y] = \sigma_t^2 + \sigma_e^2$.

3. α_c : *Not taking into account a subscale structure when* $c_s \neq 0$

If $c_s \neq 0$ and the subscale structure is not taken into account then there are *SK* discrete items. Applying Eqs. (A2.3) and (A2.7),

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

Therefore,

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

4 α_S : *Taking into account a subscale structure when* $c_s \neq 0$

Taking account of the subscale structure when $c_s \neq 0$ again simply involves summing the items within a subscale to give *S* higher order items. Applying Eqs. (A2.5) and (A2.7) to these *S* items gives

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

Therefore,

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