On the fractal dimension of a social measurement II

David Andrich
Murdoch University, Western Australia

Mailing address
David Andrich
Murdoch University
Murdoch 6150
Western Australia

Acknowledgements

The research reported in this paper was supported in part by an Australian Research Council Linkage grant with the Australian National Ministerial Council on Employment, Education, Training and Youth Affairs (MCEETYA) Performance Measurement and Reporting Task Force; UNESCO's International Institute for Educational Planning (IIEP), and the Australian Council for Educational Research (ACER) as Industry Partners. Murdoch University provided the opportunity for Study Leave which provided necessary concentrated time to develop the content of paper. Permission to use the data from the Australian Scholastic Aptitude Test, which is a test constructed by ACER, was given by the Curriculum Council of Western Australia. This paper has benefited from the many discussions with Irene Styles on the nature of social measurement and the impacts of level of scale. The work was also helped by a much earlier paper with Barry Sheridan reporting the drop in reliability for the same ASAT data set depending on the method of its calculation and continuing discussion with Guanzhong Luo regarding the structure of the Rasch model. The programming of Barry Sheridan and Guanzhong Luo was essential to being able to estimate the relevant parameters in the analysis using the Rasch model, as was Ida Marais's in simulating the data reported in the paper.

*Report No. 4 ARC Linkage Grant LP0454080: Maintaining Invariant Scales in State, National and International Level Assessments. D Andrich and G Luo Chief Investigators, Murdoch University.

On the fractal dimension of a social measurement: II

Abstract

Many scales in psychology, education and social measurement in general, which are constructed to measure a single variable, are nevertheless composed of subscales which measure different aspects of the variable. Although the presence of subscales captures better the complexity of a variable and increases its validity, it compromises its unidimensionality. This paper reconciles the measurement produced by a scale composed of subscales by resolving the measurement into a main variable common among all subscales and a set of mutually orthogonal variables unique to each subscale and orthogonal to the main variable. Then using the formula for Cronbach's α calculation of the reliability of traditional test theory, it derives a formula which estimates the summary value characterizing the main variable relative to the mutually orthogonal variables. It also derives formulae elaborating the interpretation of α calculated at different levels of scale in which account is taken of the multidimensionality produced by the subscales. A set of simulation studies, generated according to the Rasch model, illustrates the effectiveness in recovering the summary value of the mutually orthogonal variables, using both the raw scores and the Rasch model estimates of the persons. The concept of the *roughness*, adapted from fractal geometry and introduced in a companion paper (Andrich, 2006), is used as a metaphor for the impact of the mutually orthogonal variables on the main variable and a possible motif to represent this roughness is suggested. The advantage of such an approach to imperfect unidimensionality, inherent in the design structure of an instrument, is that the focus remains on the main variable to be measured. Data from an Australian Scholastic Aptitude Test (ASAT), which are analyzed illustratively in the companion paper, are reanalyzed to illustrate the interpretation of the formulae that are derived in the paper.

Key words: psychological measurement, social measurement, dimensionality, fractal dimension

On the fractal dimension of a social measurement: II

1. Introduction

Most scales constructed in psychology, education and social measurement in general, are intended to characterize more or less of some construct. Such a scale is generally said to be *unidimensional.* However, many scales are developed from an analysis of aspects of a construct, sometimes at more than one level, with a subscale identified with each aspect. In order to capture some broader aspects of the construct, these subscales are not conceptualized as perfectly, but as imperfectly, related. These different aspects capture the complexity of the construct and the presence of subscales which measure each aspect increases the validity of the scale beyond the validity that could be achieved if only one aspect were measured. An example of such a scale, analyzed illustratively in this paper, is an Australian Scholastic Aptitude Test (ASAT). It is composed of 100 multiple choice items, where these items were deliberately developed so that half of them cover mathematics and the natural sciences, and half of them the humanities and social sciences, where 26, 24, 27 and 23 items specifically assessed each of mathematics, natural science, humanities and social science respectively. The items were further grouped by reading stems, where from 3 to 8 dichotomously scored multiple choice items can be answered by reading one stem. The structure of the ASAT is shown in Figure 1.

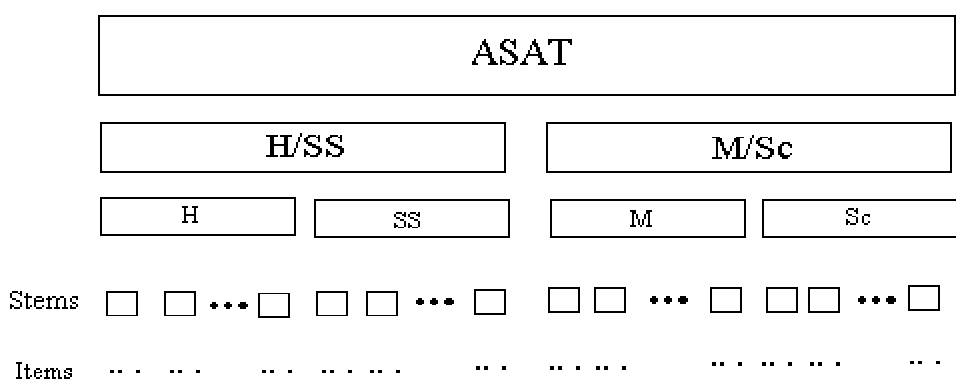

Figure 1 The subscale structure of the ASAT

In both modern and traditional test theory there are methods for identifying various violations of unidimensionality with prescriptions as to what might be done to deal with such violations. In general, the perspective is that multidimensionality contaminates the scale somehow. In circumstances where scales are constructed by a composition of subscales in order to increase validity, however, it is inevitable that there is some multidimensionality. Reducing the validity by reducing multidimensionality is a manifestation of a well known potential trade-off in scale construction, that between two required properties of scales, internal consistency reliability and external validity. Thus some multidimensionality may be seen as a positive property of the scale rather than evidence of some fault. In that case it might be useful to be able to summarize efficiently the magnitude of the main dimension of the scale relative to the dimensions of the subscales.

This paper does so in a particular way and is a companion paper to Andrich (2006) which uses the concepts of fractal geometry and the unidimensional Rasch model (Rasch, 1960/1980) to characterize the fractal dimension of a social measurement. The idea of a fractal dimension, introduced by Mandelbrot (1983), is concerned with the concept of *roughness* of a measurement. The concept of roughness seems eminently suitable for social measurement in the case of a scale composed of subscales of the kind described above.

The key observation that makes possible the calculation of a fractal dimension of a social measurement is that Rasch analyses of the same data which do and do not take account of a subscale structure, provide different units of scale and different standard deviations of the person distribution. However, when transformed linearly into the same units in the presence of a subscale structure, the standard deviations of the distribution of measurements from the different analyses are not themselves changed linearly. The non-linearity of the transformations leads to the calculation of a fractal dimension which, in the presence of multidimensionality, is a non integer greater than 1. The greater the

fractal dimension, the rougher the measurement. In the case of a fractal dimension of 1, the measurement is smooth, or in the terminology of dimensionality, unidimensional.

The derivations in this paper rest on a complementary observation regarding the standard deviation of the person distribution and the standard error of measurement from analyses at different levels of scale. In particular, when there is a subscale structure and the data are analyzed at different levels of scale, then the traditional internal consistency reliability index, defined as ratio of the true variance of the person distribution and the sum of the true plus error variances, also changes.

To study the way the traditional reliability index varies depending on the level of analysis, this paper begins by considering a scale composed of subscales, and resolves each such subscale linearly into two variables - the first is common among all subscales and characterizes the main variable; the second is unique to each subscale and is therefore mutually orthogonal with the main variable and with the variables of all other subscales. This conceptualization of roughness of a single dimension in terms of mutually orthogonal dimensions is also compatible with conceptualizations of fractal geometry where the roughness of a *surface* can be studied with orthogonal intersecting *planes* (Mandelbrot, 2002, p. 16).

The topic of dimensionality and how to deal with it in social measurement has, of course, a long history, ranging from factor analysis of the either the raw data matrix or variables that have been constructed to be relatively unidimensional (Thurstone, 1947; Harman, 1968) to the application of multidimensional item response theory models, including multidimensional versions of the Rasch model (Briggs & Wilson 2003). The distinctive feature of the approach in this and the companion paper is that the focus remains on the main variable measured by a scale with the magnitude of the mutually orthogonal variables of the subscales indicating the degree of variation, conceptualized as roughness, from unidimensionality.

Formalizing the observation of the change in reliability depending on the level of analysis of the same data, the paper derives a formula which estimates the summary value characterizing the relative magnitude of the variables mutually orthogonal to the main variable.

The derivation of the required formula which rests on the change in the reliability of traditional test theory (TTT) when analyzed at different levels of scale, is based on Cronbach's α (Cronbach, 1951). The paper summarizes the derivation of α from first principles in the usual case and then extends it to the case of a different level of scale. The advantage of Cronbach's α over other calculations of reliability for the purposes of this paper is that it can be calculated from a single administration of a scale and is general in the sense that it can be used with dichotomous and polytomous (ordered response category) items, or a combination of these with different maximum scores for different items. Both features are exploited in this paper. In addition, the value of Cronbach's α is sensitive to violations of unidimensionality, which again is relevant to this paper. This sensitivity is also the reason that α continues to receive attention, e.g. Komaroff (1997) and Rae (2006).

In TTT, items are scored with successive integers beginning with 0, and then the total score of a person on the items is taken to characterize the person responses. However, in the derivation of the Cronbach's α, it is assumed these variables, that is, the score on each item and therefore of the total score, are continuous. In deriving the equation referred to above, and in the first instance, this assumption of continuity is also made.

To illustrate the operation of the formula, however, data are simulated in which the responses to the items are discrete, as in the practical applications of the traditional test theory. The discretization of responses from latent continuous variables into dichotomous responses is carried out by postulating that the error component in the response to each item is distributed independently and homogeneously across items according the double exponential distribution (Yellott, 1977). The discretization is therefore effectively carried out using the Rasch model for dichotomous responses. The

Rasch model is particularly suitable for this purpose because the total score of a person on a scale is the key statistic in characterizing a person in both the traditional and Rasch theories - in the former by assumption, in the latter as a consequence of the model. The discretization of the responses according to the Rasch model also permits connections to be made to results of the companion paper.

The rest of the paper is structured as follows. Section 2 summarizes the derivation of Cronbach's α and the formula which estimates the summary impact of the mutually orthogonal variables relative to the main variable; Section 3 describes the simulation process connecting the Rasch model to the TTT model and provides the results of these simulations; Section 4 shows the analysis of the ASAT data. Section 5 makes more explicit connections to fractal geometry, suggests a graphical metaphor of a fractal dimension of a social measurement composed of subscales, and considers some further interpretations of the reliability formulae when subscales are taken into account Section 6 is a discussion.

2. Cronbach's α and traditional test theory

The calculation of Cronbach's α within TTT here is circumscribed by the material essential to the paper. Comprehensive summaries of TTT can be found in Gulliksen (1950), Lord and Novick (1968), Thorndike (2004).

Let the observed score of person *n* on item *i* be $x_{ni}, x_{ni} \in \{0,1\}$ and let the total score of the person on the scale be denoted by $y_n = \sum_{i=1}^{I} x_{ni}$. The development is in terms of dichotomously scored items, but the development is general. Further let

$$y_n = \tau_n + e_n \qquad (1)$$

where τ_n is the true score and e_n is the error of person *n* on the scale. As indicated above, τ_n and e_n are continuous variables, making y_n also continuous. Across a population of persons, the error is taken to be distributed normally and uncorrelated with the true scores:

$$e \approx N(0,\sigma_e^2),\ \tau \approx N(\mu,\sigma_\tau^2),\ COV[\tau,e] = 0, \tag{2}$$

where $COV[\tau,e]$ is the covariance between the true score and the error. From Eqs. (1) and (2), the total observed score variance is given by

$$V[y] = V[\tau] + V[e] + 2COV[\tau,e] = V[\tau] + V[e], \tag{3}$$

and in the value notation which will also be used in the paper,

$$\sigma_y^2 = \sigma_\tau^2 + \sigma_e^2, \tag{4}$$

where $\sigma_y^2 = V[y]$; $\sigma_\tau^2 = V[\tau]$; and $\sigma_e^2 = V[e]$.

The traditional reliability, notated for convenience immediately as α, is defined by the ratio

$$\alpha = \frac{\sigma_\tau^2}{\sigma_y^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2}, \tag{5}$$

which is clearly the *proportion* of the true score variance among persons relative to the total observed score variance.

2.1 Construction and calculation of Cronbach's α .

In calculating α , it is assumed that the response x_{ni} of each person n to each item i can also be resolved into the sum of the true score and an error according to

$$x_{ni} = \tau_n + \varepsilon_{ni} \tag{6}$$

with the analogous properties of Eq. (2) of

$$\varepsilon_{ni} \approx N(0,\sigma_i^2),\ \tau \approx N(\mu,\sigma_\tau^2),\ COV[\tau,\varepsilon_i] = 0, \tag{7}$$

where σ_i^2 is the error variance associated with an item and is assumed homogeneous across person/item engagements.

Taking two items i and j, from Eqs. (6) and (7)

$$V[x_i] = V[\tau] + V[\varepsilon_i] = \sigma_\tau^2 + \sigma_i^2, \tag{8a}$$

$$V[x_j] = V[\tau] + V[\varepsilon_j] = \sigma_\tau^2 + \sigma_j^2, \tag{8b}$$

and $$COV[x_i,x_j] = COV[\tau + \varepsilon_i, \tau + \varepsilon_j] = \sigma_\tau^2. \tag{9}$$

Now consider the variance of the total score across persons in terms of items:

$$V[y] = V[\sum_{i=1}^{I} x_i] = \sum_{i=1}^{I} V[x_i] + 2\sum_{i=1}^{I}\sum_{j>i}^{I} COV[x_i,x_j], \tag{10}$$

that is,

$$\begin{aligned} V[y] &= \sum_{i=1}^{I} V[\tau + \varepsilon_i] + 2\sum_{i=1}^{I} \sum_{j>i}^{I} COV[\tau + \varepsilon_i, \tau + \varepsilon_j] \\ &= \sum_{i=1}^{I} V[\tau] + \sum_{i=1}^{I} V[\varepsilon_i] + 2\sum_{i=1}^{I} \sum_{j>i}^{I} V[\tau], \end{aligned} \tag{11}$$

or in the alternate notation,

$$\begin{aligned} \sigma_y^2 &= \sum_{i=1}^{I} \sigma_\tau^2 + \sum_{i=1}^{I} \sigma_i^2 + 2\sum_{i=1}^{I} \sum_{j>i}^{I} \sigma_\tau^2 \\ &= I\sigma_\tau^2 + I\sigma_i^2 + I(I-1)\sigma_\tau^2, \end{aligned} \tag{12}$$

where $I(I-1)$ is the number of off diagonal elements in the matrix of covariances among items. Recall that σ_i^2 is assumed homogeneous across items where the subscript *i* is retained for clarity of exposition and therefore gives $\sum_{i=1}^{I} \sigma_i^2 = I\sigma_i^2$.

It is instructive to consider a tabular construction of the matrix of the last line of Eq. (12). This is shown in Table 1 in which the *I* variance terms $V[x_i] = \sigma_\tau^2 + \sigma_i^2$ form the diagonal terms, and the $I(I-1)$ covariance terms $COV[x_i, x_j] = \sigma_\tau^2$ form the off diagonal terms.

Table 1 Variance covariance matrix of observed item scores

$COV[x_i, x_j]$	1	2	.	j	.	I–1	I
1	$\sigma_\tau^2 + \sigma_1^2$	σ_τ^2	.	σ_τ^2		σ_τ^2	σ_τ^2
2	σ_τ^2	$\sigma_\tau^2 + \sigma_2^2$	.	.	.	.	σ_τ^2
.	.	.	.	.	.	.	.
i	σ_τ^2	.	.	.	.	.	σ_τ^2
.	.	.	.	.	.	.	.
I–1	σ_τ^2	.	.	.	.	$\sigma_\tau^2 + \sigma_{I-1}^2$	σ_τ^2
I	σ_τ^2	σ_τ^2	.	σ_τ^2	.	σ_τ^2	$\sigma_\tau^2 + \sigma_I^2$

To construct α , form the ratio

$$
\begin{aligned}
\frac{V[y] - \sum_{i=1}^{I} V[x_i]}{V[y]} &= \frac{V[\sum_{i=1}^{I} x_i] - \sum_{i=1}^{I} V[x_i]}{V[y]} \\
&= \frac{\sigma_\tau^2 + I\sigma_i^2 + I(I-1)\sigma_\tau^2 - \sigma_\tau^2 - (\sigma_\tau^2 + I\sigma_i^2)}{\sigma_\tau^2 + I\sigma_i^2 + I(I-1)\sigma_\tau^2} \\
&= \frac{I(I-1)\sigma_\tau^2}{\sigma_y^2} \\
&= \frac{I(I-1)\sigma_\tau^2}{I\sigma_\tau^2 + I\sigma_i^2 + I(I-1)\sigma_\tau^2} \\
&= \frac{I(I-1)\sigma_\tau^2}{I^2\sigma_\tau^2 + I\sigma_i^2} \\
&= \frac{I}{I}\frac{(I-1)\sigma_\tau^2}{I\sigma_\tau^2 + \sigma_i^2} \\
&= \frac{I-1}{I}\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_i^2 / I}. \qquad (13)
\end{aligned}
$$

Multiplying both sides of Eq. (13) by $I / I - 1$ gives the well known formula

$$\alpha = \frac{I}{I-1}\left(\frac{V[y]-\sum_{i=1}^{I}V[x_i]}{V[y]}\right) = \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma_i^2/I} = \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma_e^2} \tag{14}$$

where

$$\sigma_i^2/I = \sigma_e^2 \tag{15}$$

and where σ_e^2 is defined in Eq.(4).

Eq. (15) shows that the factor $I/(I-1)$ is not a correction for bias, but an inherent part of the formula for the index α. It also shows how the error $\sigma_i^2/I = \sigma_e^2$ reduces in size as the number of items increases and therefore how, for a fixed inter-item correlation, α increases correspondingly and can become much larger than this inter-item correlation. The relevance of this relationship becomes apparent again in Section 5.2 where an analogous equation in terms of subscales appears as the ratio of the true score variance and the total systematic variance.

2.2 Formalizing subscales of a scale

As indicated in the introduction, this paper considers scales composed of identifiable subscales. Let

$$\beta_{ns} = \tau_n + c_s\tau_{ns}, \tag{16}$$

where $c_s > 0$, τ_n is the common true score for person *n* among subscales and is the same variable as in Eq. (1), and τ_{ns} is the score on the *distinct* aspect characterized by subscale *s* and is uncorrelated with τ_n, that is, $COV[\tau_s, \tau] = 0$. Therefore, τ_n is the value of the

main, common, variable among subscales, and τ_{ns} is the variable unique to each subscale. The value c_s characterizes the magnitude of the variable characterized of subscale *s* relative to the common variable among subscales. The common variable τ might be considered a *higher order* variable relative to the variables τ_s of the subscales $s = 1,2,...,S$. Further, variables of the subscales are considered to be mutually uncorrelated, that is $COV[\tau_s, \tau_t] = 0$ for all subscales *s* and *t*. Table 2 shows this subscale design.

Table 2 Summary of subscale design

Subscales					
Items	1	2	.	.	*S*
1	$\beta_{n1} = \tau_n + c_1\tau_{n1}$	$\beta_{n2} = \tau_n + c_2\tau_{n2}$	.	.	$\beta_{nS} = \tau_n + c_S\tau_{nS}$
2	$\beta_{n1} = \tau_n + c_1\tau_{n1}$	$\beta_{n2} = \tau_n + c_2\tau_{n2}$	.	.	$\beta_{nS} = \tau_n + c_S\tau_{nS}$
.	.	.	.	.	.
.	.	.	.	.	.
K	$\beta_{n1} = \tau_n + c_1\tau_{n1}$	$\beta_{n2} = \tau_n + c_2\tau_{n2}$	.	.	$\beta_{nS} = \tau_n + c_S\tau_{nS}$

Because each subscale is composed of a distinct variable τ_{ns} as well as the common variable τ_n, the correlation among the subscales is *not* 1. However, because it has the common variable τ_n, and depending on the size of c_s, the correlation will generally be greater than 0. Within the above constraints, any correlation among any two subscales of a scale can be specified by setting

$$V[\tau] = V[\tau_s] = \sigma^2 \tag{17}$$

for all subscales *s* and *t*. This constraint does not lose any generality in the relationship between the true score variable and the mutually orthogonal variables of the subscales. In the derivations, the distinct notation and identities of $V[\tau]$ and $V[\tau_s]$ are maintained

for clarity of exposition, but the numerical equality is applied where is relevant to the interpretation of particular formulae.

Further modifications of the mean and variance of the variables measured by the scale can be generated by transforming β_{ns} according to

$$B_{ns} = a_s + b_s \beta_{ns} . \tag{18}$$

However, because such linear transformations do not affect the ratios of the covariances and the correlations, they are not relevant to this paper and are not considered in further derivations.

Appendix 1 shows that the theoretical, latent correlation ρ_{st} between two items in different subscales *s* and *t* from the construction in Table 2 is given by

$$\rho_{st} = \frac{1}{\sqrt{1 + c_s^2]}\sqrt{1 + c_t^2}} \quad . \tag{19}$$

Assuming that $c_s = c_t = c$ gives

$$\rho_{st} = \frac{1}{1 + c^2} . \tag{20}$$

Clearly, the larger the value of c*,* the smaller the correlation between two subscales: e.g. if $c = 0$, then $\rho_{st} = 1$, and if $c = 10$, then $\rho_{st} = 0.00999...$.

Appendix 1 shows for completeness that ρ_{st} is also the correlation between two subscales even if the number of items is different in the subscales. There are formulae where the difference between the numbers of items among subscales needs to be taken into account and these are shown in the paper.

We now derive a formula which can be used to estimate c. This is done by recalculating α in the presence of a subscale structure.

2.3 Accounting for subscales in calculating Cronbach's α .

Taking account of the subscale structure in calculating α involves taking each subscale as a higher order item whose score is the sum of the scores of the lower level items in that subscale. All the assumptions of TTT listed above are maintained at this level of analysis.

Thus, let

$$x_{nis} = \beta_{ns} + \varepsilon_{nis} = \tau_n + c_s \tau_{ns} + \varepsilon_{nis} \tag{21}$$

be the observed score of person *n* on item *i* of subscale *s*, where ε_{nis} is the error component of item *i* of subscales *s* for person *n*, and $V[\varepsilon_{si}] = \sigma_i^2$ is assumed to be homogeneous within items of a subscale and among items of different subscales, and is therefore not subscripted by *s*.

Let there be *K* items per subscale, and *S* subscales. This constraint of an equal number of items per subscale is relaxed later, but because the focus is first on the general principles rather than variations, for simplicity of exposition, equal numbers of items in each subscale are used.

Let

$$y_{ns} = \sum_{i=1}^{K_s} x_{nis} \tag{22}$$

be the total score of a person *n* on subscale *s*. For completeness and consistency of notation

$$y_n = \sum_{s=1}^{S} y_{ns} = \sum_{s=1}^{S} \sum_{i=1}^{K_s} x_{nis} \tag{23}$$

continues to be the total score of person *n* on the whole scale.

In calculating α , consider that there are *S* items each with a maximum score of *K*, rather than that there are *SK* discrete items each with a maximum score of 1.

Table 3 elaborates Table 2 and summarizes the variances and covariances among the items within a subscale and between subscales. Appendix 2 shows their derivations.

Table 3 Construction of the covariance matrix with subscales. Each entry in a cell represents one covariance between the respective elements

Subscale		1	2	…	*S*
		1,2,………*K*	1,2,……… *K*		1,2,……… *K*
1	1				
	2				
	·	$\sigma_\tau^2 + c_1^2\sigma_1^2$	σ_τ^2		σ_τ^2
	k				
2	1				
	2				
	·	σ_τ^2	$\sigma_\tau^2 + c_2^2\sigma_2^2$		σ_τ^2
	k				
S	1				
	2				
	·	σ_τ^2	σ_τ^2		$\sigma_\tau^2 + c_S^2\sigma_S^2$
	k				

It is evident from Table 3 that the covariance between two items within a subscale is greater than that between two items from different subscales by the amount $c_s^2\sigma_s^2$ where $\sigma_s^2 = V[\tau_s]$, and although postulated to be homogeneous across subscales, the subscript s is retained for purposes of exposition.

To develop the implications of a subscale structure on the value of coefficient α, the following four combinations are considered in calculating α: taking and not taking into account the presence of subscales, and for each case considering $c = 0$ and $c > 0$.

Table 4 summarizes the values of α for each combination together with the notation that is used for each, and shows explicitly the values of the numerators for each of the combinations. Because the denominator for the relevant comparisons across the rows of

Table 4 is the same, it is not expanded here, although it is considered again in Section 5.2. The derivations of these expressions are shown in Appendix 3. In these derivations, and to simplify the expressions while still retaining the sense of the effects and in order to provide a single summary value for the mutually orthogonal variables, it continues to be postulated that $c_s = c$ for all s, that is, that the covariances among all pairs of subscales are the same. Again, the subscript s is used to emphasize that it is a property of the subscales in the formulae. The notation for each of the values of α is shown in Table 4, where the subscripts c and 0 respectively indicate $c > 0$, and $c = 0$, and the superscripts (a) and (*) respectively indicate that the subscale structure is and is not taken into account. α_0 has no superscript as it is the standard case of $c = 0$ with no subscale structure, but the subscript 0 is used for emphasis and consistency.

Table 4 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
$c = 0$: standard case - subscales have a correlation of 1 .	$\alpha_0 = \dfrac{S^2K^2\sigma_\tau^2}{V[y]}$	$\alpha_0^{(a)} = \dfrac{S^2K^2\sigma_\tau^2}{V[y]}$	No effect
$c > 0$: subscales have correlation less than 1.	$\alpha_c^{(*)} = \dfrac{S^2K^2\sigma_\tau^2 + \dfrac{S^2K^2(K-1)c_s^2\sigma_s^2}{SK-1}}{V[y]}$	$\alpha_c^{(a)} = \dfrac{S^2K^2\sigma_\tau^2}{V[y]}$	Reduction of α

It is evident from the first row of Table 4 that α has the same value when $c = 0$, whether or not the subscale structure is taken into account, and from the second row when $c > 0$, that α is different when the structure is taken into account compared to when it is not.

In particular, when $c > 0$, $\alpha_c^{(*)}$ is greater than $\alpha_c^{(a)}$. However, when $c > 0$, unidimensionality is violated. Thus the results in Table 4 explain why the value of α does not indicate whether a scale measures a unidimensional variable or not, but instead provides the value of the reliability *on the assumption of unidimensionality*. This inflation of the reliability in the presence of a multidimensional subscale structure provides the basis for the main result and interpretations in this paper.

2.4 Recovering the value c and the correlation among subscales

To recover the summary value c we take the ratio of $\alpha_c^{(*)}$ and $\alpha_c^{(a)}$ from Table 4. In the simplification, the relationship $\sigma_\tau^2 = \sigma_s^2$ is applied in $\alpha_c^{(*)}$ in the numerator of Eq. (24).

$$\begin{aligned}\frac{\alpha_c^{(*)}}{\alpha_c^{(a)}} &= \frac{(S^2K^2\sigma_\tau^2 + \dfrac{S^2K^2(K-1)c^2}{SK-1}\sigma_\tau^2)/V[y_s]}{(S^2K^2\sigma_\tau^2)V[y_s]} \\ &= \frac{S^2K^2 + \dfrac{S^2K^2(K-1)c^2}{SK-1}}{S^2K^2} \\ &= 1 + \frac{(K-1)c^2}{SK-1}.\end{aligned} \tag{24}$$

Therefore

$$\frac{\alpha_{sn}}{\alpha_{st}} - 1 = \frac{(K-1)c^2}{SK-1}, \tag{25}$$

from which

$$c^2 = \left(\frac{SK - 1}{K - 1}\right)\left(\frac{\alpha_{sn}}{\alpha_{st}} - 1\right), \tag{26}$$

and which in turn implies

$$\rho_{st} = \frac{1}{1 + c^2} \ . \tag{27}$$

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

$$c^2 = \frac{\left(\frac{\alpha_{sn}}{\alpha_{st}}\right)\left(\frac{S}{S-1}\right)\left(\frac{(\sum_{s=1}^{S} k_s) - 1)}{(\sum_{s=1}^{S} k_s)}\right)\left(\sum_{s=1}^{S}\underbrace{\sum_{t=1}^{S}}_{s \neq t} k_s k_t\right) - \left(\sum_{s=1}^{S}\underbrace{\sum_{t=1}^{S}}_{s \neq t} k_s k_t\right) - \left(\sum_{s=1}^{S} k_s(k_s - 1)\right)}{\left(\sum_{s=1}^{S} k_s(k_s - 1)\right)} . \tag{28}$$

An important feature of Eqs. (26) and (28) is that the variance of the variables, σ_τ^2 and σ_s^2, together with the variances of the errors σ_i^2 and σ_e^2 are eliminated in the ratios. Thus the calculation of c, and therefore the correlation among the subscales, is obtained independently of the total variance and independently of the error variance. Although there is no space to pursue this relationship, the correlation between two variables is equivalent to a correlation corrected for attenuation due to error (Guilford, 1965). In the case of a mutual correlation among more than two variables, it is a generalization of the correction for attenuation.

The above development shows explicitly that applying the formula for α to subscales in the presence of a subscale structure gives a smaller value for α than if that structure is not taken into account. Clearly, substantial simplifications are made in the assumptions in the derivations, for example, that the correlations among items within each subscale

are homogeneous, that the variables distinctive to the subscales have a mutual correlation of 0, and that the contribution of each subscale to the overall scale through the same value of c is the same. However, these simplifications are compatible with the simplifications in the original derivation of α. The point is that when a scale is composed of defined subscales, it is possible to confirm the operational effect of the subscales by calculating α in the two ways described and applying formula (26) or (28) to obtain a summary overall latent correlation amongst the subscales independent of error. Even if the simplifying assumptions are not quite correct, as they inevitably will not be, the resultant single values c and the correlation among the subscales may be taken as the effective single *summary* value, much like the average of a distribution of numbers.

3. Simulation of data as discrete responses

As indicated in the *Introduction*, in deriving α in TTT, it is assumed that the variables are continuous, that is that y_n in $y_n = \tau_n + e_n$ of Eq. (1), x_{ni} in $x_{ni} = \tau_n + \varepsilon_{ni}$ of Eq. (6) and x_{nis} in $x_{nis} = \tau_n + c_s\tau_{ns} + \varepsilon_{nis}$ of Eq. (21) are continuous. However, in practice, the responses of persons to items are discrete, often dichotomous values 0 or 1 as in this study, and therefore the total scores y_n are also discrete.

Therefore, to illustrate the application of the formula derived above in this practical situation, data are simulated so that the true and error scores, taken as *latent*, are continuous, but the manifest response variables x_{ni}, x_{nis} and therefore y_n are discrete. This is done by formulating the trait parameters τ_n and τ_{ns}, and the error distributions ε_{ni} and ε_{nis} as continuous, but then, in conjunction with taking account of the location of each item, converting the continuous response to a discrete response. For convenience of exposition, let

$$z_{ni} = \mu_{ni} + \varepsilon_{ni} = \tau_n - \delta_i + \varepsilon_{ni} \qquad (29)$$

and

$$\begin{aligned} z_{sni} &= \mu_{nis} + \varepsilon_{nis} \\ &= \beta_{ns} - \delta_{is} + \varepsilon_{nis} \\ &= \tau_n + c_s\tau_{ns} - \delta_{is} + \varepsilon_{nis}\ , \end{aligned} \tag{30}$$

where δ_i is the location of item i on the same scale as τ_n of person n and δ_{is} is the location of item i in subscale s.

3.1 Constructing discrete responses using the dichotomous Rasch model.

Instead of the normal distribution as in TTT, let the errors ε_{ni} be distributed according to the double exponential distribution with mean 0 and variance $\pi^2/3$=3.289868 (and standard deviation 1.813799) (Bock and Jones, 1968). This distribution is normal for all practical purposes (Bock and Jones, 1968). Figure 2 shows this distribution of errors around the difference $\mu_{ni} = \tau_n - \delta_i = 1$ between the ability τ of a person and the difficulty δ of an item.

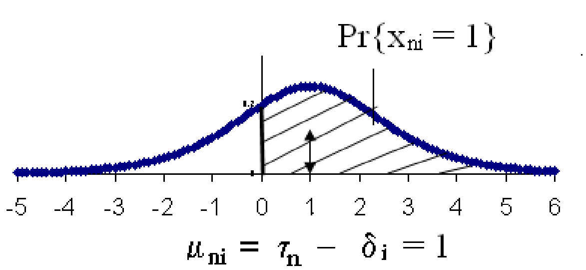

Figure 2 The double exponential distribution with mean 1 and variance $\pi^2/3$. The shaded area shows $P\{x_{ni} = 1\}$.

Formally,

$$\begin{aligned} P\{x_{ni} = 1 \mid \tau_n, \delta_i\} &= \int_0^\infty \frac{\exp(z_{ni} - \mu_{ni})}{(1 + \exp(z_{nj} - \mu_{ni}))^2} dz_{ni} \\ &= \frac{\exp \mu_{ni}}{1 + \exp \mu_{ni}} = \frac{\exp(\tau_n - \delta_i)}{1 + \exp(\tau_n - \delta_i)}, \end{aligned} \tag{31}$$

and $P\{x_{ni} = 0 \mid \tau_n, \delta_i\} = 1 - P\{x_{ni} = 1\}$.

Eq. (31) is the familiar Rasch model for dichotomous responses.

In the case that the responses are made discrete in the presence of subscales, then

$$\begin{aligned} P\{x_{nis} = 1 \mid \tau_n, \delta_i\} &= \int_0^\infty \frac{\exp(z_{nis} - \mu_{nis})}{(1 + \exp(z_{nis} - \mu_{nis}))^2} dz_{sni} \\ &= \frac{\exp \mu_{nis}}{1 + \exp \mu_{nis}} = \frac{\exp(\tau_n + c_s \tau_{ns} - \delta_{is})}{1 + \exp(\tau_n + c_s \tau_{ns} - \delta_{is})} \end{aligned} \tag{32}$$

3.1 Design of the simulation studies

Eq. (26) which estimates c^2 and equations leading to it are algebraic and therefore tautologous. The properties of this estimator, e.g. bias, are not studied in this paper algebraically and are left for another occasion. In addition, discretizing the continuous variables will affect the estimates. The degree of this effect is examined in preliminary simulations on the quality of the recovery of c^2 according to Eq. (26).

Two design prototypes are taken in these illustrative studies. The first has two subscales and involves three studies with respectively 15, 20 and 30 dichotomous items in each subscale; the second has four subscales with respectively 10, 15, and 20 dichotomous items each. For each of these, the correlation between subscales is specified to be 1 ($c = 0.0$), 0.8 ($c = 0.5$), and 0.5 ($c = 1$). The value of $c = 0.0$ provides a base line for the

studying the application of the formula in the case where there is no subscale structure. In the companion paper it was indicated that if $c = 0.0$, then analyzing the responses in terms of subscales has no effect. Results in row 1 of Table 4 show that this assertion is correct algebraically, and the simulation studies demonstrate that it is very much correct empirically.

Each particular simulation was repeated five times with different random seeds, and the results show the summary statistics from these replications. Essentially, the value of c_s^2 is recovered in each study, averaged over five simulations, and then converted to the correlation according to Eq. (27). In each simulation, 1000 persons were distributed normally with mean 0, standard deviation 2, and the items were uniformly distributed between -4.0 and 4.0 inclusive. The design of the studies and the recovery of the relevant values are shown in Table 4. However, before considering these values, another approach to calculating α which arises from the Rasch model is described briefly.

3.2 Calculation and application of the reliability index from Rasch model estimates

The generic definition of traditional reliability, notated in this paper immediately by α, is given by Eq. (5) and repeated for convenience

$$\alpha = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} . \tag{33}$$

It is possible to estimate the parameters of Eq. (33) using parameter estimates recovered from a Rasch model analysis. In particular, let

$$\hat{\tau}_n = \tau_n + \varepsilon_n \tag{34}$$

be the resolution of the Rasch ability estimate $\hat{\tau}_n$ into a component τ_n which is free of measurement error, and ε_n be the error of measurement with the usual assumption that the error is uncorrelated with the estimate in a population. Then

$$\sigma_{\hat{\tau}}^2 = \sigma_{\tau}^2 + \sigma_e^2 \tag{35}$$

and $$\sigma_{\tau}^2 = \sigma_{\hat{\tau}}^2 - \sigma_e^2 \quad . \tag{36}$$

An estimate of $\sigma_{\hat{\tau}}^2$ in Eq. (36) is readily calculated from

$$\hat{\sigma}_{\hat{\tau}}^2 = \sum_{n-1}^{N} (\hat{\tau} - \bar{\hat{\tau}})^2 /(N-1) \tag{37}$$

where $\hat{\tau}$ is the estimate of the person abilities obtained from an analysis of data using the Rasch model.

In addition, with each person estimate $\hat{\tau}_n$, an estimate $\hat{\sigma}_{ne}^2$ of the variance of the error of this estimate is available. The average of this variance,

$$\hat{\sigma}_e^2 = \sum_{n=1}^{N} \hat{\sigma}_{ne}^2 / N \, , \tag{38}$$

can be taken as an estimate of σ_e^2 in Eq. (36).

Then an estimate of α from the Rasch person location estimates, notated α_R in this paper, can be obtained by inserting the values obtained from Eqs. (37) and (38) into Eq. (39):

$$\hat{\alpha}_R = \frac{\hat{\sigma}_\tau^2}{\hat{\sigma}_\tau^2 + \sigma_e^2} = \frac{\hat{\sigma}_{\hat{t}}^2 - \hat{\sigma}_e^2}{\hat{\sigma}_{\hat{t}}^2}. \qquad (39)$$

Given that the data were generated using the Rasch model in the step which converted the continuous response distributions into discrete responses, it is relevant to show the values of both α and α_R in the simulation studies and the recovered correlations among the subscales using both α and α_R. This is done in Tables 5 and 6 for the respective simulation designs. A potential advantage of the reliability formulation from the Rasch estimates relative to the calculating Cronbach's α from raw scores is their linearization of raw scores, especially those that are close to the maximum and minimum scores. It also makes possible connections between the results in this paper and those in the companion paper which is based on analyzing the ASAT data using the Rasch model.

In applying the Rasch model to data when the subscales are taken into account in which new items are formed as sums of the original items within each scale, it is necessary to use the generalization of the model for two ordered categories to the model with more than two ordered categories. This model, and its application are summarized in the companion paper (Andrich, 2006), and are therefore not repeated in this paper.

The software used to obtain the Rasch estimates for Eq.(39), RUMM2020 (Andrich, Sheridan and Luo, 2005), uses weighted likelihood estimates (Warm,1989) for the person parameters, given estimates of the item parameters. These have substantially reduced bias relative to maximum likelihood estimates. The item parameters themselves are estimated using a pairwise conditional method of estimation (Andrich and Luo, 2003), in which the person parameters are eliminated. Estimates of the person and item parameters are not reported in this paper, as the effectiveness of estimating item and person parameters in the Rasch model has been well established in the literature.

3.3 Results of simulation studies

Tables 5 and 6 show the design of the simulations together with the recovery of the constant c^2 and the corresponding correlation ρ_{st} for each of the studies based on values of α and α_R using Eq. (26). Also shown in the tables are the theoretical values of ρ_{st} and the actual values r_{st} in the simulations in order better to separate random simulation effects from estimation effects. The standard deviations of the values of c^2 over the 5 replications are also shown.

It is evident in Tables 5 and 6 that there is not a noticeable difference between the recovery of c^2 from α compared to α_R and that the recovery of the values of c_s^2 are excellent. Since the study is not directly on this comparison, and is carried out for completeness, it will not be commented on further. In addition, because the point of the paper is to illustrate an approach to characterizing measurement in the presence of a subscale structure, and the estimates are evidently excellent with little evidence of misleading interpretations, in order that they do not distract from this point, no statistical tests of the quality of the recovery of c^2 relative to the generating values or comparisons between α and α_R are carried out in this paper. These important issues if the approach is taken up, are left for other studies. However, because connections are made in the paper to the Rasch analyses of the companion paper, for efficiency, results only from α_R of Rasch analyses of ASAT in the next section are provided. The results from the direct calculation of α are commensurate with those from α_R.

Table 5 Design 1 of the simulation studies and recovery of the constant c^2 and the correlation among subscales

					Items per subscale	
Design 1:	Theoretical	Actual		Study 1 15	Study 2 20	Study 3 30
2 subscales	c^2 (ρ_{st})	c^2 (r_{st})	From	$\hat{c}^2$ ($\hat{r}_{st}$) sd $\hat{c}^2$	$\hat{c}^2$ ($\hat{r}_{st}$) sd $\hat{c}^2$	$\hat{c}^2$ ($\hat{r}_{st}$) sd $\hat{c}^2$
	0.000 (1.000)	0.000 (1.000)	α	-0.024 (1.025) sd 0.005	-0.017 (1.017) sd 0.004	-0.011 (1.012) sd 0.002
			α_R	-0.001 (1.001) sd 0.006	0.000 (1.000) sd 0.006	-0.000 (1.000) sd 0.002
	0.250 (0.800)	0.253 (0.798)	α	0.225 (0.817) sd 0.020	0.231 (0.813) sd 0.034	0.243 (0.805) sd 0.026
			α_R	0.269 (0.788) sd 0.020	0.275 (0.784) sd 0.033	0.282 (0.780) sd 0.023
	1.000 (0.500)	1.025 (0.494)	α	0.932 (0.518) sd 0.076	1.000 (0.500) sd 0.153	1.000 (0.501) sd 0.116
			α_R	1.052 (0.487) sd 0.066	1.193 (0.456) sd 0.177	1.197 (0.455) sd 0.151

$\tau \approx N(0,4)$; $-4.0 \leq \delta_{is} \leq 4.0$

Table 6 Design 2 of the simulation studies and recovery of the constant c^2 and the correlation among subscales

Design 2:	Theoretical	Actual		Study 1 10	Study 2 15	Study 3 20
4 subscales	c^2	c^2	From	$\hat{c}^2$	$\hat{c}^2$	$\hat{c}^2$
	(ρ_{st})	(r_{st})		($\hat{r}_{st}$)	($\hat{r}_{st}$)	($\hat{r}_{st}$)
				sd $\hat{c}_s^2$	sd $\hat{c}_s^2$	sd $\hat{c}_s^2$
	0.000	0.000	α	-0.040	-0.022	-0.017
	(1.000)	(1.000)		(1.042)	(1.022)	(1.017)
				sd 0.004	sd 0.003	sd 0.004
			α_R	-0.001	0.001	0.000
				(1.001)	(1.000)	(1.000)
				sd 0.004	sd 0.006	sd 0.006
	0.250	0.251	α	0.195	0.233	0.231
	(0.800)	(0.799)		(0.837)	(0.811)	(0.813)
				sd 0.018	sd 0.018	sd 0.034
			α_R	0.248	0.270	0.275
				(0.801)	(0.788)	(0.784)
				sd 0.013	sd 0.011	sd 0.033
	1.000	1.006	α	0.934	1.000	1.000
	(0.500)	(0.499)		(0.517)	(0.500)	(0.500)
				sd 0.082	sd 0.060	sd 0.153
			α_R	1.056	1.072	1.193
				(0.486)	(0.483)	(0.456)
				sd 0.084	sd 0.075	sd 0.177

$\tau \approx N(0,4)$; $-4.0 \le \delta_{is} \le 4.0$

A second important observation from Tables 5 and 6 is that the recovery of c^2 in the case that $c = 0$ (ρ_{st} =1) is excellent. In particular, from α_R the greatest difference from a theoretical value of 1 is 0.001. This provides the reference point for the calculations when the value of c is not 0. In some cases the value of c^2 is slightly negative which is a manifestation of random variation in the estimates around 0. The important feature of this result is that when c = 0 and theoretically there is no multidimensionality, it is recovered well by the formula of Eq. (26).

The conclusion indicates that $\hat{c} > 0$ implies multidimensionality. Therefore, it would be helpful to have a sampling distribution for c. Such a distribution has not been derived but Monte Carlo studies, such as those used in this paper, are available in any important research in which such a conclusion is important.

4 Analysis of the ASAT

As indicated in the *Introduction*, the items of ASAT are constructed at different levels of scale. They can be analyzed in terms of subscales in terms of these levels, in particular: (i) simply as 100 distinct multiple choice items; (ii) as 17 polytomous items where the score on each item is the sum of the scores on the multiple choice items associated with each reading stem; (iii) as four polytomous items where the score on each item is the sum of the scores of the stems in each of the discipline areas of Mathematics, Science, Humanities and Social Science; (iv) as two polytomous items where the score on one item is the sum of the scores of the items in the discipline areas of Mathematics and Science, and the score on the other item is the sum of the scores of the items in the discipline area of Humanities and Social Science.

Relative to each successive analysis, it is possible to apply Eqs. (28) or its special case Eq. (26) in order to estimate the corresponding value of c^2. Table 7 shows the results for each of the successive pairs of analyses. In addition, for completeness, the case of two subscales formed from the 100 items directly is also shown.

The value of $\hat{c}$ is relatively large in each successive analysis. It is evident that the reliability drops with each successive analysis and therefore that there is noticeable multidimensionality as measured by this value. This is the case also when two subscales are formed directly from the 100 items rather than in successive steps, and indeed drops further, suggesting that such an analysis takes account of more dependencies.

Table 7 Estimated values of c^2, c and r for the ASAT data

	Run as 100 distinct items	Run as 17 reading stem items	$\hat{c}^2$	$\hat{c}$	r	A_s
α_R	0.922	0.893	0.542	0.736	0.649	0.966
	Run 17 reading stems	Run as 4 discipline items				
α_R	0.893	0.827	0.307	0.554	0.765	0.926
	Run as 4 discipline items	Run as 2 combined discipline items				
α_R	0.827	0.732	0.391	0.626	0.719	0.836
	Run as 100 distinct items	Run as 2 combined discipline items	0.922			
α_R	0.922	0.732	0.524	0.724	0.656	0.792

Further interpretation of this Table is made after the concept of the fractal dimension is considered in the next Section. Table 7 also shows a value denoted by A_s. This value is derived in the Section 5.3, and is used to interpret further the results in Table 7.

5 A possible conceptual rendition of a fractal dimension

In the companion paper, the concept of a fractal dimension was introduced through the changes in length of a coastline and the geometrical Koch motif at different levels of scale. The analogous property to length in calculating the fractal dimension of the ASAT test was the standard deviation of the distribution of persons relative to changes in the units of scale for analyses conducted at different levels of scale, that is, forming higher order items from the sums of the items in the subscales, and analyzing the data in terms of these higher order items. It was indicated in that paper that it was not obvious how to

construct a motif analogous to the Koch motif that might carry the idea of the roughness of a measurement. The formulation of this paper seems to permit a construction and therefore might help carry the idea of a fractal dimension for social measurement more concretely.

5.1 The systematic variance of a variable

The systematic variance of a variable is briefly summarized here as it is used first to help construct a motif to capture the roughness of a variable, and in the next section to further interpret the different calculations of α .

From Eqs. (16) and (23), and again assuming $k_1 = k_2 = ... = k_s = K$ for convenience,

$$y_n = \sum_{s=1}^{S} y_{ns} = \sum_{s=1}^{S} \sum_{i=1}^{K} x_{nis} = \sum_{s=1}^{S} \sum_{i=1}^{K} (\beta_{ns} + \varepsilon_{nis}) = \sum_{s=1}^{S} \sum_{i=1}^{K} (\tau_n + c_s \tau_{ns} + \varepsilon_{nis}) . \quad (40)$$

That is,

$$y_n = \sum_{s=1}^{S} \sum_{i=1}^{K} (\tau_n + c_s \tau_{ns}) + \sum_{s=1}^{S} \sum_{i=1}^{K} \varepsilon_{nis} \quad (41)$$

Let

$$T_n = \sum_{s=1}^{S} \sum_{k=1}^{K} (\tau_n + c_s \tau_{ns}) = \sum_{s=1}^{S} \sum_{k=1}^{K} \tau_n + \sum_{s=1}^{S} \sum_{k=1}^{K} c_s \tau_{ns} . \quad (42)$$

Then T_n is the sum of the common true score and the unique components of the variable for person *n*, and being free of error, it can be described as the *systematic* variance of the variable y_n among persons.

Eq. (42) may be simplified to

$$T_n = SK\tau_n + \sum_{s=1}^{S} Kc_s\tau_{ns} . \tag{43}$$

Then the total systematic variance is given by

$$V[T] = S^2K^2V[\tau] + \sum_{s=1}^{S} K^2c^2V[\tau_s] = S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2 . \tag{44}$$

Eq. (44) shows that the total systematic variance increases according to two additive components, one composed of the true score variance common to the subscales, the other the variances unique to each component and orthogonal to the true score variance and proportional to c^2.

Invoking the simplifying relationship $\sigma_\tau^2 = \sigma_s^2 = \sigma^2$ of the magnitudes of the variances, give the total systematic variance as

$$V[T] = (S^2K^2 + SK^2c^2)\sigma^2 . \tag{45}$$

5.2 A possible motif for a fractal dimension of a social measurement involving subscales

In the companion paper, the standard deviations of the scale and the person distributions respectively were considered to correspond to the unit and length in physical measurement. In the construction of subscales, the formulation according to the value of the systematic component of the variable in the response of a person to one item is given in Eq. (16) by $\beta_{ns} = \tau_n + c_s\tau_{ns}$. Therefore,

$$V[\beta_s] = \sigma_\tau^2 + c^2\sigma_s^2 . \tag{47}$$

Applying $\sigma_\tau^2 = \sigma_s^2 = \sigma^2$ of the magnitudes of the variances again gives

$$V[\beta_s] = (1 + c_s^2)\sigma^2. \tag{48}$$

The variance in Eq. (48) can be represented as a right triangle with sides of length σ and $c\sigma$ respectively with the variances then being the squares on the sides of the right triangle. Such a triangle is shown in Figure 3.

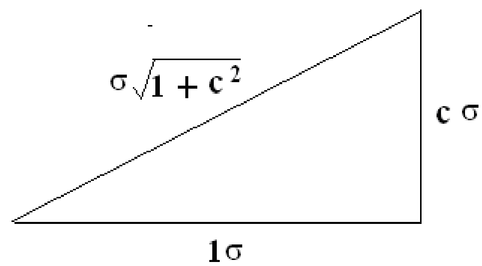

Figure 3 The right triangle with sides representing relative standard deviations of an item of a subscale.

Figure 3 shows that the total variance of the common and orthogonal variables of a responses to an item are greater, namely $(1 + c^2)\sigma^2$, rather than σ^2 as is the case when $c = 0$. This suggests an approach to a diagrammatic rendition of roughness, more or less analogous to the Koch motif shown in the companion paper.

Thus consider the following construction. From Eq.(45),

$$V[T] = (S^2K^2 + SK^2c^2)\sigma^2 = SK(SK\sigma^2 + Kc^2\sigma^2). \quad (46)$$

Therefore a component of variance of $(SK\sigma^2 + Kc^2\sigma^2)$ can be associated with each item, there being *SK* items in total. Then a right triangle whose sides are respectively $\sqrt{SK}\sigma$ and $(\sqrt{K}c\sigma)$ can be formed and its sum of squares is the variance component $(SK\sigma^2 + Kc^2\sigma^2)$. By juxtaposing these triangles *SK* times, a motif can be drawn. To illustrate this, consider a special case of $S = 2$ subscales and K= 4 items per subscale. Then

$$\sqrt{SK}\sigma = \sqrt{8}\sigma = 2\sqrt{2}\sigma \text{ and } \sqrt{K}c\sigma = 2c\sigma . \quad (49)$$

A suggested motif is shown in Figure 4. The upper graph shows a case where $c = 0$ and the variance from A to B on the upper line and B to A on the lower line is simply the sum of the squares of $2\sqrt{2}\sigma$ on each segment of the lines giving

$$V[T] = S^2K^2\sigma^2 = SK(\sqrt{SK}\sigma)^2 = 8(2\sqrt{2})^2\sigma^2 = 64\sigma^2.$$

The lower graph shows how the variance from A to B and B to A is greater in the presence of $c \neq 0$, very much like a rough coastline length increases as the unit of measurement is reduced:

$$\begin{aligned} V[T] &= SK((\sqrt{SK}\sigma)^2 + (\sqrt{K}c\sigma)^2) = 8((2\sqrt{2})^2\sigma^2 + (2c\sigma)^2) \\ &= 8(8\sigma^2 + 4c^2\sigma^2) = 64\sigma^2 + 32c^2\sigma^2 > 64\sigma^2. \end{aligned}$$

It suggests a rougher measurement than the first graph, which is effectively a straight line rather than a jagged one.

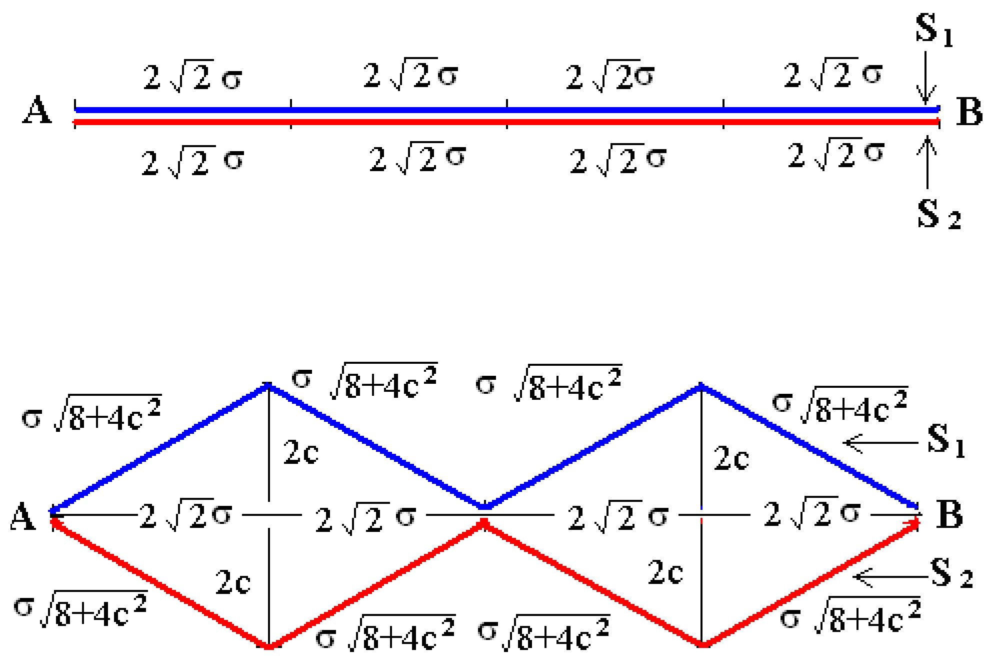

Figure 4. A suggested motif for two subscales with four items per subscale

The details of the graph of Figure 4 in itself are not as important as visualizing how the *systematic* part of the variance increases in the presence of subscales when the items are analyzed as if the scales were not present, that is, if they are analyzed as discrete items.

5.3 The proportion of common variance relative to the systematic variance

The above interpretation of the true score variance and the variance orthogonal to it that gives rise to the multidimensionality suggests a closer analysis of the calculations of α and the relationship between $\alpha_c^{(*)}$ and $\alpha_c^{(a)}$ defined in Table 4.

First note that

$$\begin{aligned}\alpha_c^{(*)} &= (S^2K^2\sigma_\tau^2 + \frac{S^2K^2(K-1)c_s^2\sigma_s^2}{SK-1})/V[y] \\ &= (S^2K^2\sigma_\tau^2 + \frac{S^2K^2(1-1/K)c_s^2\sigma_s^2}{(S-1/K)})/V[y].\end{aligned} \tag{50}$$

Then, since $\lim_{K\to\infty}(1-1/K)/(S-1/K)=1/S$, and $(1-1/K)/(S-1/K)<1/S$ for finite K,

$$\begin{aligned}\alpha_c^{(*)} &\leq (S^2K^2\sigma_\tau^2 + \frac{S^2K^2c^2\sigma_s^2}{S})/V[y] \\ &= (S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2)/V[y],\end{aligned} \tag{51}$$

where from Eq. (41)

$$V[y] = S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2 + \sigma_e^2 \tag{52}$$

is the total variance. That is

$$\alpha_c^{(*)} \leq \frac{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2}{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2 + \sigma_e^2} \tag{53}$$

Second, it is now evident both from the denominator of Eq. (51) and from Eq. (5), that the numerator of Eq. (53) is the systematic variance (composed of both the true variance and the variances of the subscales orthogonal to it), indicating that $\alpha_c^{(*)}$ is the limit of the ratio of the systematic variance relative to the total variance as the number of items per subscale increases, that is, it is a lower bound of the proportion of systematic variance relative to the total variance.

We may now relate $\alpha_c^{(a)}$, which is the ratio of the common true score variance to the total systematic and error variance, to $\alpha_c^{(*)}$ as follows:

$$\begin{aligned}\alpha_c^{(a)} &= \frac{S^2K^2\sigma_\tau^2}{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2 + \sigma_e^2} \\ &= \frac{S^2K^2\sigma_\tau^2}{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2}\frac{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2}{S^2K^2\sigma_\tau^2 + SK^2c^2\sigma_s^2 + \sigma_e^2},\end{aligned} \tag{54}$$

that is

$$\alpha_c^{(a)} \le A_S \alpha_c^{(*)} \tag{55}$$

where $$A_S = \frac{S^2 K^2 \sigma_\tau^2}{S^2 K^2 \sigma_\tau^2 + SK^2 c^2 \sigma_s^2} \tag{56}$$

is the proportion of true variance relative to the systematic variance. Thus A_S is constructed much like α, except that it pertains to the systematic variance only. In summary, $\alpha_c^{(a)}$, which accounts for the multidimensionality among subscales and is the proportion of the true variance relative to the total variance (which is the sum of the true common, unique, and error variances), is bounded by the *product* of the proportion of true common variance relative to the systematic variance and the proportion of the systematic variance relative to the total variance.

Eq. (56) may be simplified further by applying the relationship of the magnitudes of the variances $\sigma_\tau^2 = \sigma_s^2 = \sigma^2$ to give

$$\begin{aligned} A_S &= \frac{S^2 K^2}{S^2 K^2 + SK^2 c^2} \\ &= \frac{1}{1 + c^2 / S}. \end{aligned} \tag{57}$$

Clearly, as the value of c increases, so the proportion of the true variance relative to the systematic variance increases. However, for a fixed value of c, we also observe that

$A_S = \dfrac{1}{1 + c^2 / S} > \rho_{st} = \dfrac{1}{1 + c^2}$ for any $S > 1$, where ρ_{st} is the summary mutual correlation among pairs of subscales given by Eq. (20).

In the case that it is chosen to analyze the data without accounting for multidimensionality, then A_S gives the value of the relative contribution of the true and

unique variance rather than ρ_{st} which is the correlation among subscales, even in the case that $S = 2$.

The reason for this effect is that, as can be seen from Eqs. (44) and (45), for the addition of each subscale, the true variance component of the total systematic variance $V[T]$ increases quadratically as a function of the number of subscales S (i.e. $S^2K^2\sigma^2$), while the unique variance component increases linearly ($SK^2c^2\sigma^2$). The effect is the same as that which produces the traditional α reliability of a test to be greater than the mutual summary correlation among the items shown in Eq. (14).

This effect justifies considering whether or not the addition of a subscale contributes to the main true score variance at a faster rate than it contributes unique variance. Thus if a subscale is removed, and the reliability increases, then the subscale is contributing more unique than common variance and should be removed or modified. Again, this is analogous to the case of adding items in the usual case where no subscales are present. In this case, if the α reliability increases when an item is removed, then it implies that it is contributing more error variance (which likely includes unidentifiable unique variance) than it is contributing common true score variance.

In the case of a different number of k_s items per subscale, A_s generalizes to

$$A_s = \frac{1}{1 + c^2(\sum_{s=1}^{S} k_s^2)/(\sum_{s=1}^{S} k_s)^2}. \qquad (58)$$

where the relationships summarized above continue to hold, in particular, that as the number of subscales increases, then the proportion of true variance relative the systematic variance increases.

5.4 Interpretation of A_S in relation to the ASAT analysis

Table 5 included an estimate of A_S for the successive analyses of the ASAT. The first level of analysis took account only of any multidimensionality arising from the items belonging to reading stems, and the systematic variance included both the common true score variance across all items and any unique variance of the disciplines. This latter value, 0.966, is very high showing that of the systematic variance, most was left as true score common variance. However the proportion of systematic variance itself relative to the total variance dropped to 0.893 from 0.922.

The second level of analysis, when the reading stems were grouped into each of four disciplines, took account of the unique variance arising from each of the four disciplines as well as that arising from the reading stems. In this case the proportion of the systematic variance that is true variance dropped marginally to 0.926. The proportion of the systematic variance relative to the total variance dropped to 0.827.

The third level of analysis, when the discipline areas were grouped into pairs at the level of social science/humanities and mathematics/science, took account of the unique variance arising from each of the pairs of disciplines as well as that arising from the individual disciplines and the reading stems. In this case the proportion of the systematic variance dropped more noticeably to 0.836, while the proportion of the systematic variance itself dropped to 0.732.

The last row of Table 5 shows the proportion of systematic variance that is true variance when the 50 items from each of the humanities and social sciences and mathematics and natural sciences are considered individual items but placed into two subscales. Here the drop in A_S is slightly greater, to 0.792, indicating that a greater degree of multidimensionality is accounted for among the items in this construction of subscales than when successive scales are summed.

The most interesting conclusion, however, is that even in this case where the maximum unique variance was accounted for, giving a relatively low α_R reliability of 0.732 compared to the original high and relatively inflated value of 0.922, the proportion of the remaining systematic variance which is true score variance is 0.792, close to 0.8.

6 Discussion

This section considers some implications of the perspective developed in the paper, and there are many more that are beyond the scope of the present paper. However, an important qualification is made. In the development of the paper, and in the analyses, only one of the two major violations of the Rasch model and traditional test theory was considered, that of multidimensionality. The other one, that of response dependence which in the language of modern test theory is termed local independence, was not considered. If the response structure violates local independence and so forms subscales with locally dependent items in this way, then analyses according to the subscales described above will show similar effects, that is a value of c will be estimated which is greater than 0. Clearly, the context and the nature of the subscales, the construction of the items, and their theoretical relationships need to be considered in making any interpretation. In addition, real data sets are likely to have both violations operating more or less at the same time. However, whichever violations of the model assumptions are present, analyses according to subscales, and an estimated value of c, ρ_{st} and A_s can be helpful as a backdrop to finer grained understanding of the response matrix.

6.1 Implications for tests of fit.

It is clear *a-priori* that a test composed of items with a subscale structure, and depending on the value of c, is likely to have items not fit the Rasch model if the items are analyzed as discrete independent items. It might be suggested that the two parameter model which accounts for item discrimination (Birnbaum, 1968), might be more suitable. However, that implies that differences in discrimination are only a property of the item rather than a property of the relationships among items (Humphry, 2005). Clearly, this is not the case.

One possible approach to handling the subscale structure better than simply analyzing all items independently and making inferences on the fit of the items individually when they may be violating the model because of the subscale structure, is first to check items for gross violations at the level of the scale as a whole as well as to check them at a finer level within each subscale. Then whether to use results from an analysis accounting for the subscale structure or one not accounting for this structure, would depend on the purpose. In making decisions, two consequences follow. First, no one particular value for an index of fit should determine a decision as to whether an item is worth retaining or not. All statistical information available, together with the operation of the other items and the purpose of the testing, will need to be taken into account. Second, the level of scale needs to be taken into account if any linking or other relationships are to be formed with other tests.

6.2 Power of the tests of fit

The perspective above draws together aspects of traditional test theory and Rasch measurement theory more closely than might have been considered possible.
The value of the traditional reliability is also a strong indicator of the power of tests of fit in the Rasch model - that is the greater the spread of the persons, the greater the overall power of the test fit and should be considered in conjunction with all the information available in making decisions about items. Thus the power of the test of fit when the subscale structure is ignored will generally be greater in detecting misfit than when the subscale structure is taken into account. Of course, at the level of scale of this case, there is a genuine greater likelihood that the data do misfit the model.

6.3 Roughness of scale and perfection of fit

In the companion paper, the idea of roughness was introduced as a perspective on recognizing a subscale structure. It was suggested that a degree of roughness might make a variable more valid than if it were perfectly smooth. Two final points are made for

completeness in relation to this perspective. First, the perspective taken in these papers in relation to subscales is simply a generalization of the case of having more items in a scale where effectively every item forms its own subscale. Every additional item that conforms to the scale adds to the precision of measurement of the variable common to all items and it also introduces some element unique to the item. This unique element will, in both TTT and Rasch measurement theory, be formally absorbed into the error. However, the item with its unique element will add to the complexity and therefore the validity of the test. The extreme case where each item does not add some systematic variance unique to the item, though potentially different in location from other items, would render the item entirely redundant in substance. If the error were reduced by adding many such items, then the attenuation paradox of TTT in which the reliability increases to the point that the validity is only that of one item, would also be approached. The discussion in this paper and the companion paper extends this feature of the contribution of the substantive uniqueness of each item to the level of each subscale with some of its implications.

One related difference that arises from applying the Rasch model and using location estimates corresponding to raw scores rather than using only the raw scores, is that the variance of the person locations corresponding to each total score changes depending on the level of analysis. This is the case even after accounting for unit of scale so that the actual true score variance of the person estimates is smaller when the subscale structure is taken into account compared to when it is not. This is the main subject of the companion paper (Andrich, 2006). In TTT, the observed raw scores at the level of the test remain unchanged irrespective of how they are analyzed. The interesting consequence that seems to be implied is that any correlation of the Rasch model location estimates when accounting for a subscale structure may produce a lower correlation with a predictor variable than an analysis that ignores the subscale structure. This interpretation is consistent with the comments on the complexity and validity of variables. Clearly, some optimum between perfect unidimensionality and multidimensionality is required in each case, but it suggests that the focus on unidimensionality in tests of model fit might need to be tempered by the context.

Second, unlike many other tests of significance, it needs to be recognized that in item response theory in general, and in Rasch measurement in particular, the significance tests imply deviations from perfection, not deviations from randomness, as many statistical tests are formulated. As empirical tests are inevitably imperfect from a strictly statistical perspective, and given that this imperfection will inevitably be shown given enough power, for example a large enough sample size or large enough distance between persons and items, or both, decisions on tests of fit must also take account of the full range of evidence that is available in any analysis and the use to which the test data will be put. It is suggested that the use of mechanical tests of fit for exclusion of items should not be carried out. Of course, the reasoning in making decisions regarding items, for example to modify, discard, or retain them, should be made clear in any publication.

In summary, this paper considers a particular slant on issues that are commonly grappled with in educational, psychological and social measurement in general, that of the structure of subscales of a scale, by bringing together the traditional and Rasch measurement theories and relating them to the concept of roughness found in the field of fractal geometry. This slant has many further implications beyond those summarized in this and the companion paper which may be worth studying.

Appendix 1

The latent correlation among two items in different subscales independent of error

From Eq. (16)

$\beta_{ns} = \tau_n + c_s \tau_{ns}$ for subscale s and therefore $\beta_{nt} = \tau_n + c_t \tau_{nt}$ for subscale t.

Therefore

= = ; = = .

Further, because

.

= = = .

Therefore the correlation ρ_{st} between two items from different subscales is given by

$$\rho_{st} = \frac{COV[\beta_s, \beta_t]}{\sqrt{V[\beta_s]}\sqrt{V[\beta_t]}} = \frac{COV[\beta_s, \beta_t]}{\sqrt{V[\tau + c_s \tau_s]}\sqrt{V[\tau + c_t \tau_t]}}$$

$$= \frac{V[\tau]}{\sqrt{V[\tau] + c_s^2 V[\tau_s]}\sqrt{V[\tau] + c_t^2 V[\tau_t]}}.$$

Given $V[\tau] = V[\tau_s] = V[\tau_t]$,

$$\rho_{st} = \frac{1}{\sqrt{1 + c_s^2}\sqrt{1 + c_t^2}}.$$

For $c_s = c_t = c$, $\rho_{st} = \dfrac{1}{1 + c^2}$.

For K_s, K_t items in subscales *s* and *t* respectively, the total latent scores, independent of error, for person *n* are given respectively by

$$K_s\beta_{ns} = K_s\tau_n + K_s c_s \tau_{ns}\,;\;\; K_t\beta_{nt} = K_t\tau_n + K_t c_t \tau_{nt}\,.$$

Therefore,

$$V[K_s\beta_{ns}] = K_s^2 V[\tau] + K_s^2 c_s^2 V[\tau_s];\;\; V[K_t\beta_{nt}] = K_t^2 V[\tau] + K_t^2 c_t^2 V[\tau_t]$$

and

$$COV[K_s\beta_{ns}, K_t\beta_{nt}]] = K_s K_t V[\tau]\,.$$

Therefore the correlation ρ'_{st} between two subscales is given by

$$\rho'_{st} = \frac{K_s K_t V[\tau]}{\sqrt{K_s^2 V[\tau] + K_s^2 c_s^2 V[\tau_s]}\sqrt{K_t^2 V[\tau] + K_t^2 c_t^2 V[\tau_t]}}$$
$$= \frac{K_s K_t V[\tau]}{K_s K_t \sqrt{V[\tau] + c_s^2 V[\tau_s]}\sqrt{V[\tau] + c_t^2 V[\tau_t]}}$$

and again with $V[\tau] = V[\tau_s] = V[\tau_t]$,

$$\rho'_{st} = \frac{1}{\sqrt{1 + c_s^2}\sqrt{1 + c_t^2}} = \rho_{st}$$

which is identical to the correlation between two items from different subscales.

Appendix 2

The covariance among items within the same subscale and between different subscales

From Eq. (21) $x_{ni} = \tau_n + c_s\tau_{ns} + \varepsilon_{nis}$, and (i) $COV[\tau,\tau_s] = 0$, (ii) $COV[\varepsilon_i,\tau_s] = 0$, and for any subscale s , (i) $V[\tau] = V[\tau_s] = V[\tau_t]$, and (ii) $COV[\tau_t,\tau_s] = 0$ for subscales s and t.

Then

$$V[x_{ni}] = V[\tau] + c_s^2V[\tau_s] + V[\varepsilon_i].=\sigma_\tau^2 + c_s^2\sigma_\tau^2 + \sigma_i^2 = (1+c_s^2)\sigma_\tau^2 + \sigma_i^2.$$

For two items *i* and *j* in the same subscale s,

$$\begin{aligned} COV[x_{is}, x_{js}] &= COV[\tau + c_s\tau_s + \varepsilon_{is}, \tau + c_s\tau_s + \varepsilon_{is}] \\ &= V[\tau + c_s\tau_s] = V[\tau] + c_s^2V[\tau_s] = V[\tau] + c_s^2V[\tau] = (1+c_s^2)\sigma_\tau^2. \end{aligned}$$

If $c_s = 0$, then $COV[x_{is}, x_{js}] = \sigma_\tau^2$.

For two items in different subscales *s* and *t*,

$$\begin{aligned} COV[x_{is}, x_{jt}] &= COV[\tau + c_s\tau_s + \varepsilon_{is}, \tau + c_t\tau_t + \varepsilon_{it}] \\ &= V[\tau] = \sigma_\tau^2. \end{aligned}$$

Appendix 3

Calculation of α under four different conditions

1. α_0: Not taking into account a subscale structure when there is no structure (standard case)

The case when there is no subscale structure is the standard case. However, for comparing the formulae, consider that there are *S* sets of *K* items each. In this case, the covariance among all *SK* items is the same. The only modification in Eq. (14) for the calculation of α in this case is that the total number of items is *SK*. For simplicity, the denominator is not expanded and is the same whether the subscale structure is taken into account or not. The formula is derived from first principles rather than by substitution of expressions for completeness in anticipation of deriving the formulae for the other cases.

Therefore, the *numerator* in Eq. (14) takes the form

$$\begin{aligned}
N[\alpha_0] &= \frac{SK}{SK-1}\left(V[y] - \sum_{i=1}^{SK} V[x_i]\right) = \frac{SK}{SK-1}\left(V[\sum_{i=1}^{SK} x_i] - \sum_{i=1}^{SK} V[x_i]\right) \\
&= \frac{SK}{SK-1}\left(\sum_{i=1}^{SK} V[x_i] + SK(SK-1)COV[x_i,x_j] - \sum_{i=1}^{SK} V[x_i]\right) \\
&= \frac{SK}{SK-1}\left(SK\sigma_\tau^2 + SK\sigma_i^2 + SK(SK-1)\sigma_\tau^2 - SK\sigma_\tau^2 - SK\sigma_i^2\right) \\
&= \frac{SK}{SK-1} SK(SK-1)\sigma_\tau^2 \\
&= S^2K^2\sigma_\tau^2.
\end{aligned}$$

Therefore

$$\alpha_0 = S^2K^2\sigma_\tau^2 / V[y].$$

2 $\alpha_0^{(a)}$ Taking into account a subscale structure when there is no subscale structure

Summing items within each subscale gives *S* items.

From Eq. (23),

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

Accordingly

$$\begin{aligned} COV[y_s, y_t] &= COV[\sum_{i=1}^{K} x_{is}, \sum_{j=1}^{K} x_{jt}] \\ &= COV[\sum_{i=1}^{K} (\tau + c_s\tau_s + \varepsilon_{si}), \sum_{i=1}^{K} (\tau + c_t\tau_t + \varepsilon_{ti})] \end{aligned}.$$

Because $c_s = c_t = c = 0$,

$$\begin{aligned} COV[y_s, y_t] &= COV[\sum_{i=1}^{K} (\tau + \varepsilon_s), \sum_{i=1}^{K} (\tau + \varepsilon_t)] \\ &= K^2 V[\tau] = K^2 \sigma_\tau^2. \end{aligned}$$

Now calculate α when there are *S* items rather than *SK* items under the above conditions.

The *numerator* in Eq.(14) takes the form

$$\begin{aligned} N[\alpha_0^{(a)}] &= \frac{S}{S-1}(V[y] - \sum_{i=1}^{I} V[y_s]) \\ &= \frac{S}{S-1}(\sum_{i=1}^{I} V[y_s] + S(S-1)COV[y_s, y_t] - \sum_{i=1}^{I} V[y_s]) \\ &= \frac{S}{S-1}(S(S-1)COV[y_s, y_t]) \\ &= \frac{S}{S-1} S(S-1)K^2\sigma_\tau^2 \\ &= S^2K^2\sigma_\tau^2 \end{aligned}$$

Therefore

$\alpha_0^{(a)} = S^2K^2\sigma_\tau^2 / V[y]$.

3 α_c^*: Not taking into account a subscale structure when items within each subscale do have a structure.

The case considered here is where α is calculated when there is a subscale structure, that is, $c \neq 0$, but no account is taken of this structure. That is, it is assumed that there are *SK* independent items.

Taking that the number of items is *SK*, gives

$$N[\alpha_c^{(*)}] = \frac{SK}{SK-1}(V[y] - \sum_{s=1}^{S}\sum_{i=1}^{K} V[x_i]) = \frac{SK}{SK-1}(V[\sum_{s}^{S}\sum_{i=1}^{K} x_i] - \sum_{s=1}^{S}\sum_{i=1}^{K} V[x_i])$$

$$= \frac{SK}{SK-1}\sum_{s}^{S}\sum_{i=1}^{K} V[x_i] + S(K(K-1))COV[x_i,x_j] + S(S-1)K^2 COV[x_i,x_t] - \sum_{s=1}^{S}\sum_{i=1}^{K} V[x_i])$$

$$= \frac{SK}{SK-1}(S(K(K-1))(\sigma_\tau^2 + c_s^2\sigma_\tau^2) + S(S-1)K^2\sigma_\tau^2)$$

$$= \frac{SK}{SK-1}(SK^2 - SK)\sigma_\tau^2 + (SK^2 - SK)c_s^2\sigma_\tau^2 + (S^2 - S)K^2\sigma_\tau^2).$$

That is,

$$N[\alpha_c^{(*)}] = \frac{SK}{SK-1}(SK^2 - SK + (SK^2 - SK)c_s^2 + S^2K^2 - SK^2)\sigma_\tau^2$$

$$= \frac{SK}{SK-1}(S^2K^2 - SK + (SK^2 - SK)c_s^2)\sigma_\tau^2$$

$$= \frac{SK}{SK-1}(SK(SK-1) + SK(K-1)c_s^2)\sigma_\tau^2$$

$$= (S^2K^2 + \frac{S^2K^2(K-1)c_s^2}{SK-1})\sigma_\tau^2$$

$$S^2K^2\sigma_\tau^2 + \frac{S^2K^2(K-1)c_s^2}{SK-1}\sigma_\tau^2.$$

Therefore

$$\alpha_c^{(*)} = (S^2K^2\sigma_\tau^2 + \frac{S^2K^2(K-1)c_s^2\sigma_\tau^2}{SK-1})/V[y].$$

4 $\alpha_c^{(a)}$: Taking into account a subscale structure when items within each subscale do have a structure.

Summing items within each subscale gives S items. Substituting directly into the numerator for Eq. (14) for α gives

$$\begin{aligned}
N[\alpha_c^{(a)}] &= \frac{S}{S-1}(V[y] - \sum_{s=1}^{S} V[y_s]) = \frac{S}{S-1}(V[\sum_{s=1}^{S} y_s] - \sum_{s=1}^{S} V[y_s]) \\
&= \frac{S}{S-1}(\sum_{s=1}^{S} V[y_s] + S(S-1)COV[y_s, y_t] - \sum_{i=1}^{I} V[y_s]) \\
&= \frac{S}{S-1}(S(S-1)COV[y_s, y_t]) \\
&= \frac{S}{S-1} S(S-1)K^2\sigma_\tau^2 \\
&= S^2K^2\sigma_\tau^2
\end{aligned}$$

Therefore

$$\alpha_{st} = S^2K^2\sigma_\tau^2 / V[y].$$

References

Andrich, D. (2006). On the fractal dimension of a social measurement: I. Unpublished Paper, Social Measurement Laboratory, Murdoch University, Perth, Australia.

Andrich, D. and Luo, G. (2003) Conditional estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement 4,* 205 - 221.

Andrich, D., Sheridan, B.S. and Luo, G. (2005) RUMM2020: Rasch Unidimensional Models for Measurement. RUMM Laboratory: Perth, Western Australia.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick, *Statistical theories of mental test scores* (pp. 397 –545). Reading, Mass.: Addison-Wesley.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Bock, R.D. & Jones, L.V. (1968). *The measurement and prediction of judgment and choice.* San Francisco: Holden Day.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, *4(1),* 87-100.

Guilford, J.P. (1965). Fundamental statistics in psychology and education.(Fourth Edition) New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York: Wiley.

Harman, H.H. (1968) *Modern factor analysis*. (Second Edition) Chicago, The University of Chicago Press.

Humphry, S.M. (2005). *Maintaining a common arbitrary unit in social measurement*. Ph.D. Thesis: http://wwwlib.murdoch.edu.au/adt/browse/view/adt-MU20050830.95143

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

Lauwerier, H. (1987) *Fractals: Endlessly repeated gormetrical figures*. Princeton: Princeton University Press.

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

Mandelbrot, B.B. (1983). *The Fractal Geometry of Nature*. San Franciso**:** Freeman.
Rae, G (2006). Correcting coefficient alpha for correlated errors: Is α_k a lower bound for reliability.

Rasch, G. (1960/80). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980) with foreword and afterword by B.D. Wright, (1980). Chicago: The University of Chicago Press.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54,* 427-450.

Thurstone, L.L. (1947) Multiple Factor Analysis. Chicago, Illinois: The University of Chicago Press.

Thorndike, R. L. (2004) Measurement And Evaluation In Psychology And Education, 7th Edition, Prentice Hall:

Yellott, J.I. Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology, 15,* 109-144.