Estimation of reliability: a better alternative for Cronbach's alpha

Kimmo Vehkalahti [a,*], Simo Puntanen [b], Lauri Tarkkonen [a]

[a] *Department of Mathematics and Statistics, P.O. Box 54, FI-00014 University of Helsinki, Finland*

[b] *Department of Mathematics, Statistics & Philosophy, FI-33014 University of Tampere, Finland*

Abstract

We study the estimation of the reliability of measurement scales. The subject has been partially obscured because of poor estimators, such as Cronbach's alpha, which is the most widely applied estimator of reliability. We show that a recently proposed estimator that we are calling Tarkkonen's rho, provides a general method for the task. Examining the assumptions behind these estimators we prove that Cronbach's alpha is a special case of Tarkkonen's rho under certain one-dimensional measurement models. In practice, the conditions required for Cronbach's alpha to be a valid estimator of reliability are difficult to encounter, whereas Tarkkonen's rho is well suited for more general measurement settings. Our conclusion is that Tarkkonen's rho is a better alternative for Cronbach's alpha.

Key words: Reliability, Cronbach's alpha, Measurement error, Measurement model, Measurement scale
AMS 2000 subject classification: 62H25, 62H20, 91C05

1 Introduction

Measurement is an essential concept in science, forming the basis of all statistical research. The uncertainties of measurement are related to validity and reliability. In general, validity tells whether a measuring instrument measures

* Corresponding author.
Email address: `Kimmo.Vehkalahti@helsinki.fi` (Kimmo Vehkalahti).

what it is supposed to measure in the context in which it is applied. Obviously, validity should always be the primary concern, but we should also be interested in the accuracy of the measurements, that is, the reliability.

The reliability is defined as the ratio of the true variance to the total variance of the measurement. The true variance does not include the variance of the random measurement error. These concepts are defined in the test theory of psychometrics [1]. The concept of reliability originates from Spearman's [2,3] early work with factor analysis and measurement errors over hundred years ago.

In order to have an estimate of the reliability, an estimate of the measurement error variance is needed. This is the basic problem in the estimation of reliability, and various solutions have been suggested. In some areas, it is typical to assume that the measurement error variance is known, based on, say, earlier studies or repeated measurements (see, e.g., [4]). This may work well, e.g., in technical applications, but may not be plausible in areas such as the social sciences or the behavioral sciences, where the measurements could be far from stable.

Another solution has been based on making suitable assumptions to obtain an estimator in a form that does not include any trace of the assumptions. Nevertheless, the assumptions are implicitly present, and affect the estimates. One particular estimator has been known by many names beginning from the 1930s, but since 1951 it has been called *Cronbach's alpha* [5]. The original motivation behind it was to find quick methods for practical needs—long before the era of computers.

Despite the fact that Cronbach's alpha underestimates the reliability and may even give absurd, negative estimates [6–9], it is the most widely applied estimator of reliability. Alternative estimators based on factor analysis have been suggested [10,11], but they have not succeeded to replace Cronbach's alpha. Instead, the research has often focused on examining the properties of Cronbach's alpha without questioning its original assumptions (see a review in [12, pp. 630–635], or more recent studies, e.g., [13–16]). The motivation of estimating the reliability and using the estimates for assessing the quality of the measurements has been obscured, because the estimates have been so poor. Weiss and Davison [12] stated this in 1981: "Somewhere during the three-quarter century history of classical test theory the real purpose of reliability estimation seems to have been lost" [12, p. 633].

In this paper, we study the estimation of reliability, discussing its grounds and purposes. We focus on two estimators: 1) Cronbach's alpha [5], and 2) a new estimator proposed recently by Tarkkonen and Vehkalahti [17]. Our claim is that the new estimator provides a better alternative for Cronbach's alpha, because it is defined for more general circumstances with less assumptions, and hence it does not suffer from any serious restrictions. To prove our claim, we compare the estimators and examine their assumptions in detail.

Section 2 reviews the central concepts and presents the details of the estimators. Section 3 compares the estimators and summarizes the results. Section 4 concludes.

2 Estimation of reliability

In this section, we review the concepts behind the reliability and its estimation to present the details of the estimators under study.

The concept of reliability is defined in the test theory of psychometrics (see [1, pp. 56–61]). The measurements x are assumed to consist of unobserved true scores τ and unobserved measurement errors ε as $x = \tau + \varepsilon$, where the expectation $\mathrm{E}(\varepsilon) = 0$ and the covariance $\mathrm{cov}(\tau, \varepsilon) = 0$. It follows that the variance $\mathrm{var}(x) = \sigma_x^2 = \sigma_\tau^2 + \sigma_\varepsilon^2$.

The reliability of x is generally defined as the ratio σ_τ^2/σ_x^2, but since

$$\mathrm{cov}(x, \tau) = \mathrm{cov}(\tau + \varepsilon, \tau) = \mathrm{cov}(\tau, \tau) = \mathrm{var}(\tau) = \sigma_\tau^2,$$

it can also be seen as the squared correlation between x and τ:

$$\rho_{x\tau}^2 = \frac{[\mathrm{cov}(x, \tau)]^2}{\mathrm{var}(x)\,\mathrm{var}(\tau)} = \frac{(\sigma_\tau^2)^2}{\sigma_x^2 \sigma_\tau^2} = \frac{\sigma_\tau^2}{\sigma_x^2}.$$

The notation $\rho_{x\tau}^2$ is often used in the literature, because the true score is taken as a scalar in the test theory of psychometrics. The point in the definition of the reliability is the ratio of the variances. In order to obtain an estimate of the reliability, however, either σ_τ^2 or σ_ε^2 must be estimated. Various estimators of reliability have been suggested, but we will mainly focus on two of them: 1) Cronbach's alpha and 2) a new estimator we are suggesting to be called *Tarkkonen's rho*. Both estimators are based on the above definition of reliability.

2.1 Cronbach's alpha

We begin by summarizing the development of Cronbach's alpha. We find it useful to provide such a summary, since it helps to understand the assumptions behind this estimator and its predecessors. For more comprehensive reviews, see, e.g., [18] or [12].

2.1.1 Predecessors and assumptions

Cronbach's alpha is essentially based on the work of Kuder and Richardson [19] in the 1930s. At that time, the concept of measurement model was not always explicitly referred to in the literature, but implicitly it was the one-dimensional factor model originated by Spearman [2] in 1904. The measured items were usually dichotomous, because the calculations had to be done manually. The primary requirement for an estimator of reliability was simplicity. These conditions formed the background for the predecessor of Cronbach's alpha, the *Kuder–Richardson formula 20*, which can be derived as follows.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be p-dimensional random vectors with a $(2p \times 2p)$ covariance matrix

$$\mathrm{cov}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} := \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix},$$

where $\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{yx}$. Then, denoting symbolically,

$$\boldsymbol{\Sigma}_{xx} = \begin{pmatrix}\sigma_1^2 & \cdots & \rho_{ij}\sigma_i\sigma_j\\ \vdots & \ddots & \vdots\\ \rho_{ij}\sigma_i\sigma_j & \cdots & \sigma_p^2\end{pmatrix} \text{ and } \boldsymbol{\Sigma}_{xy} = \begin{pmatrix}\rho_{11}\sigma_1^2 & \cdots & \rho_{ij}\sigma_i\sigma_j\\ \vdots & \ddots & \vdots\\ \rho_{ij}\sigma_i\sigma_j & \cdots & \rho_{pp}\sigma_p^2\end{pmatrix},$$

where $\sigma_i^2 = \mathrm{var}(x_i)$, $\rho_{ij} = \mathrm{cor}(x_i, y_j)$, and ρ_{ii} denotes the reliability of the single item x_i. Note that

$$\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{xx} - (\boldsymbol{I}_p - \boldsymbol{\rho})\boldsymbol{\Sigma}_d,$$

where $\boldsymbol{I}_p$ is an identity matrix of order p, $\boldsymbol{\rho} = \mathrm{diag}(\rho_{11}, \ldots, \rho_{pp})$, and

$$\boldsymbol{\Sigma}_d = \mathrm{diag}(\boldsymbol{\Sigma}) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2).$$

To proceed further, it is assumed that [19, pp. 156–157]

(1) the correlations are equal, denoted by $\rho_{ij} = \rho_{xx}$ for all i, j, and that
(2) the standard deviations are equal, i.e., $\sigma_i = \sigma_j = \sigma_x$ for all i, j.

From (1) it follows that also the reliabilities of the items are assumed to be equal, i.e., $\rho_{ii} = \rho_{xx}$ for all i.

The above assumptions imply that x_i and y_j are so-called *parallel measurements* (see [1, p. 48]), that is, $\boldsymbol{x}$ and $\boldsymbol{y}$ have an intraclass correlation structure with a covariance matrix

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix} = \sigma_x^2\begin{pmatrix}\boldsymbol{\Sigma} & \rho_{xx}\mathbf{1}\mathbf{1}'\\ \rho_{xx}\mathbf{1}\mathbf{1}' & \boldsymbol{\Sigma}\end{pmatrix} := \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{pmatrix},$$

where $\boldsymbol{\Sigma} = (1-\rho_{xx})\boldsymbol{I}_p + \rho_{xx}\mathbf{1}\mathbf{1}'$ and $\mathbf{1}$ is the vector of ones.

The parallel measurements can be formulated as a parallel model [2]

$$\begin{cases} x_i = \tau_i + \varepsilon_i \\ y_j = \tau_j + \varepsilon_j \end{cases}$$

with the assumptions $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$ and $\operatorname{cov}(\tau_i, \varepsilon_i) = \operatorname{cov}(\tau_j, \varepsilon_j) = 0$. In addition, it is assumed that $\operatorname{var}(\varepsilon_i) = \operatorname{var}(\varepsilon_j)$ and that $\operatorname{cov}(\tau_i, \tau_j) = \operatorname{cov}(\tau_i, \tau_i) = \operatorname{var}(\tau_i)$. Denoting $\operatorname{var}(x) = \operatorname{var}(x_i)$ and $\operatorname{var}(\tau) = \operatorname{var}(\tau_i)$, we obtain

$$\rho_{xx} = \frac{\operatorname{cov}(\tau_i + \varepsilon_i, \tau_j + \varepsilon_j)}{\operatorname{var}(x)} = \frac{\sigma_\tau^2}{\sigma_x^2} = \rho_{x\tau}^2,$$

that is, the item reliabilities, the common correlations of the intraclass correlation structure, are in accordance with the definition of the reliability. We note that the strict assumptions of equal variances of the measurement errors (and thus equal variances of the items) are not required in the definition of the reliability, they are merely properties of the parallel model.

Now, consider two measurement scales, the unweighted sums $u = \mathbf{1}'\boldsymbol{x}$ and $v = \mathbf{1}'\boldsymbol{y}$. The variance of u is

$$\sigma_u^2 = \operatorname{var}(\mathbf{1}'\boldsymbol{x}) = \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} = p(p-1)\sigma_x^2\rho_{xx} + p\sigma_x^2, \tag{2.1}$$

and the correlation between u and v can be written, using (2.1), as

$$\begin{aligned}\rho_{uv} = \operatorname{cor}(\mathbf{1}'\boldsymbol{x}, \mathbf{1}'\boldsymbol{y}) &= \frac{\mathbf{1}'\boldsymbol{\Sigma}_{xy}\mathbf{1}}{\operatorname{var}(\mathbf{1}'\boldsymbol{x})}\\ &= \frac{\sigma_x^2\, p^2\, \rho_{xx}}{\sigma_x^2\, p\,[1+(p-1)\rho_{xx}]}\\ &= \frac{p\,\rho_{xx}}{1+(p-1)\rho_{xx}},\end{aligned} \tag{2.2}$$

which is known as the *Spearman–Brown formula*, after Spearman [20] and Brown [21] in 1910.

Solving ρ_{xx} from (2.1) yields

$$\rho_{xx} = \frac{\sigma_u^2 - p\,\sigma_x^2}{p(p-1)\sigma_x^2}, \tag{2.3}$$

and substituting it in (2.2) gives

$$\rho_{uv} = \frac{p}{p-1}\left(1 - \frac{p\,\sigma_x^2}{\sigma_u^2}\right), \tag{2.4}$$

which is known as the Kuder–Richardson formula 20 or KR-20, where 20 refers to the number of the formula in the original article [19] in 1937.

The item reliabilities ρ_{xx} had been a nuisance with the earlier methods of reliability estimation. In the derivation of KR-20 (2.4) they were hidden by algebraic manipulations, because the aim was to provide quick methods for practical needs. However, it was clear that the figures obtained would be underestimates if the strict assumptions of equal correlations and standard deviations were not met [19, p. 159].

2.1.2 Generalization with a shorter name

The principal advantages claimed for KR-20 were ease of calculation, uniqueness of estimate (compared to so-called split-half methods suggested by Spearman [20]), and conservatism. The method was also criticized, because the magnitude of the underestimate was unknown, and even negative values could be obtained. One of the critics was Cronbach, who in 1943 stated that "while conservatism has advantages in research, in this case it leads to difficulties" [22, p. 487] and that "the Kuder–Richardson formula is not desirable as an all-purpose substitute for the usual techniques" [22, p. 488].

In 1951, Cronbach presented "the more general formula" [5, p. 299]

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\sigma_{x_i}^2}{\sigma_u^2}\right), \tag{2.5}$$

and adviced that we should take it "as given, and make no assumptions regarding it" [5, p. 299]. It is easy to see that Cronbach's alpha (2.5) resembles KR-20 (2.4): only the term $p\,\sigma_x^2$ is replaced by $\sum_{i=1}^{p}\sigma_{x_i}^2$. This loosens the strict assumption of the equal variances made in the derivation of KR-20.

Exactly the same formula (2.5) had earlier been derived by Guttman [23], who showed that the formula will give the true value of the reliability only if the variances and covariances of the true scores are all equal [23, p. 274]. Novick and Lewis [7] have later named this condition *τ-equivalence*. The corresponding measurement model is more general than the parallel model, since the variances of the measurement errors (and thus the variances of the items) need not to be equal. However, the problem is that the variances or the covariances of the true scores do not appear at all in Cronbach's alpha (2.5) or KR-20 (2.4).

2.1.3 Some interpretations

Cronbach presented the formula (2.5) without a derivation in order to "proceed in the opposite direction, examining the properties of alpha and thereby arriving at an interpretation" [5, p. 299]. One interpretation of Cronbach's alpha is that it tells how uniformly the items contribute to the unweighted sum. This property, known as the *internal consistency* of the scale, depends on the correlations between the scale and the items. We show this briefly below.

Let the scale be the unweighted sum $u = \mathbf{1}'\boldsymbol{x}$ and let x_i be a single item of that scale. The covariance of the scale and the item is

$$\mathrm{cov}(u, x_i) = \mathrm{cov}(\mathbf{1}'\boldsymbol{x}, x_i) = \sum_{j=1}^{p} \sigma_{x_j x_i},$$

and hence the variance of the scale can be expressed as

$$\mathrm{var}(u) = \sigma_u^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{x_j x_i} = \sum_{i=1}^{p} \mathrm{cov}(u, x_i) = \sigma_u \sum_{i=1}^{p} \rho_{u x_i}\sigma_{x_i}, \tag{2.6}$$

where ρ_{ux_i} is the correlation between the scale and the item. Using (2.6), Cronbach's alpha becomes

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\sigma_{x_i}^2}{(\sum_{i=1}^{p}\rho_{ux_i}\sigma_{x_i})^2}\right), \tag{2.7}$$

and we can interpret that the internal consistency of the scale is improved, when the items have high positive correlations with the scale. The term $\rho_{ux_i}\sigma_{x_i}$ is called the *reliability index* (see, e.g., [1]). If the items are standardized, i.e., $\sigma_{x_i}^2 = 1$ for all i, we can rewrite (2.7) simply as

$$\alpha = \frac{p}{p-1}\left(1 - \frac{p}{(\sum_{i=1}^{p}\rho_{ux_i})^2}\right), \tag{2.8}$$

which clearly shows how Cronbach's alpha is a function of the correlations between the scale and the items. Denoting $\sum_{i=1}^{p}\rho_{ux_i}$ by S, we can infer from (2.8) that $-\infty < \alpha \le 1$, assuming that $p > 1$. In particular,

$$\begin{cases} \alpha < 0, & \text{if } 0 < S < \sqrt{p}, \\ \alpha = 0, & \text{if } S = \sqrt{p}, \\ 0 < \alpha < 1, & \text{if } \sqrt{p} < S < p, \text{ and} \\ \alpha = 1, & \text{if } S = p, \text{ i.e., when } \rho_{ux_i} = 1 \text{ for all } i. \end{cases}$$

It is also evident that

$$\lim_{S\to 0} \alpha = -\infty,$$

which corresponds to the case where $\mathrm{var}(u) = 0$.

Using matrix notation and emphasizing the scale weights in the notation, Cronbach's alpha (2.5) can be written in the form

$$\alpha(\mathbf{1}) = \frac{p}{p-1}\left(1 - \frac{\operatorname{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right),$$

where $\operatorname{tr}(\cdot)$ refers to the trace. Since $\operatorname{tr}(\boldsymbol{\Sigma}) = \mathbf{1}'\boldsymbol{\Sigma}_d\mathbf{1}$, it can also be given as

$$\alpha(\mathbf{1}) = \frac{p}{p-1}\left(\frac{\mathbf{1}'(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_d)\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) = \frac{p^2\bar{\sigma}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}},$$

where $\bar{\sigma}$ is the average item covariance (see, e.g., [24,9])

$$\bar{\sigma} = \frac{\mathbf{1}'(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_d)\mathbf{1}}{p(p-1)}.$$

Also this result is even simpler if the items are standardized [8, p. 19]. Let us denote the correlation matrix of the items by $\boldsymbol{R} = \boldsymbol{\Sigma}_d^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_d^{-1/2}$. Obviously, $\boldsymbol{R}_d = \operatorname{diag}(\boldsymbol{R}) = \boldsymbol{I}_p$ and $\mathbf{1}'\boldsymbol{R}_d\mathbf{1} = \operatorname{tr}(\boldsymbol{R}) = p$. Now,

$$\alpha(\mathbf{1}) = \frac{p}{p-1}\left(\frac{\mathbf{1}'(\boldsymbol{R} - \boldsymbol{R}_d)\mathbf{1}}{\mathbf{1}'\boldsymbol{R}\mathbf{1}}\right) = \frac{p\bar{\rho}}{1 + (p-1)\bar{\rho}},$$

i.e., when the items are standardized, Cronbach's alpha is equal to the Spearman–Brown formula (2.2), where the unknown item reliability ρ_{xx} is estimated by the average item correlation

$$\bar{\rho} = \frac{\mathbf{1}'(\boldsymbol{R} - \boldsymbol{R}_d)\mathbf{1}}{p(p-1)} = \frac{\mathbf{1}'\boldsymbol{R}\mathbf{1} - p}{p(p-1)}.$$

2.1.4 Generalization of the scale

The unweighted sum was often found too limited for practical needs, which motivated a generalization in a form $u = \boldsymbol{a}'\boldsymbol{x}$, where $\boldsymbol{a} \in \mathbb{R}^p$. The reliability of a weighted sum was considered by Mosier [24], who proposed the formula

$$r_{uu} = \frac{\boldsymbol{a}'\boldsymbol{R}^*\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{R}\boldsymbol{a}}, \qquad (2.9)$$

where $\boldsymbol{R}$ is the correlation matrix of the items and $\boldsymbol{R}^*$ is the same matrix with the item reliabilities ρ_{ii} on the diagonal, i.e., $\operatorname{diag}(\boldsymbol{R}^*) = \operatorname{diag}(\rho_{11}, \ldots, \rho_{pp})$. Despite the estimation problems caused by the item reliabilities, the formula (2.9) has been useful, e.g., in finding a scale that maximizes the reliability.

Also Cronbach's alpha has been generalized for the weighted scale in the form

$$\alpha(\boldsymbol{a}) = \frac{p}{p-1}\left(1 - \frac{\boldsymbol{a}'\boldsymbol{\Sigma}_d\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}\right),$$

but of course, replacing the unit weights by arbitrary weights easily violates the original assumptions and may lead to doubtful results. Nevertheless, the weighted form of Cronbach's alpha has been extensively applied in practice and also employed in procedures that involve maximizing the reliability (see, e.g., [25–27,8,9]). One example is *alpha factor analysis* [26], which has been criticized because of its paradoxical results [28].

2.2 Tarkkonen's rho

Recently, a new estimator of reliability has been proposed by Tarkkonen and Vehkalahti [17] as a part of their measurement framework. Since the estimator appears in the form (2.13) already in Tarkkonen's PhD thesis [29], we suggest the estimator to be called Tarkkonen's rho. In the following, we review the key concepts behind this estimator.

2.2.1 Measurement model

Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ measure a k-dimensional ($k < p$) true score $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ with a random measurement error $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$. It is assumed that $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}) = \mathbf{0}$. The measurement model is [17, p. 176]

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{\tau} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\mu}$ is the expectation of $\boldsymbol{x}$ and $\boldsymbol{B} \in \mathbb{R}^{p \times k}$ specifies the relationship between $\boldsymbol{x}$ and $\boldsymbol{\tau}$. Denoting $\mathrm{cov}(\boldsymbol{\tau}) = \boldsymbol{\Phi}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, it follows that

$$\mathrm{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{B}\boldsymbol{\tau}) + \mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{\Psi}.$$

It is assumed that $\boldsymbol{B}$ has full column rank, and that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are positive definite [17, p. 176]. We may use the triplet $\mathscr{M} = \{\boldsymbol{x}, \boldsymbol{B}\boldsymbol{\tau}, \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{\Psi}\}$ to denote the measurement model shortly.

2.2.2 Measurement scale

The variables $\boldsymbol{x}$ are used in further analyses by creating measurement scales $\boldsymbol{u} = (u_1, \ldots, u_m)'$ as linear combinations $\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{p \times m}$ is the weight matrix. There are various criteria for choosing the weights [17, pp. 177–178], but generally it is assumed that $\boldsymbol{A}$ has full column rank and $\boldsymbol{B}'\boldsymbol{a}_i \neq \mathbf{0}, i = 1, \ldots, m$, where $\boldsymbol{a}_i$ is the ith column vector of $\boldsymbol{A}$.

Essential for the estimation of reliability is the decomposition [17, p. 177]

$$\mathrm{cov}(\boldsymbol{u}) = \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A} + \boldsymbol{A}'\boldsymbol{\Psi}\boldsymbol{A}, \tag{2.10}$$

where the total variation of $\boldsymbol{u}$ is split in two parts: 1) the variation generated by the true scores, and 2) the variation generated by the measurement errors.

2.2.3 Reliability

The variances of the measurement scales are decomposed similarly, by extracting the diagonals of the matrices in (2.10). Hence we have a diagonal matrix

$$(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})_d = \text{diag}(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}) = (\boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A})_d + (\boldsymbol{A}'\boldsymbol{\Psi}\boldsymbol{A})_d, \tag{2.11}$$

and Tarkkonen's rho is obtained, according to the definition of reliability, by dividing the variance expressions in (2.11) by another. In the general case of m scales, Tarkkonen's rho is a diagonal $m \times m$ matrix [17, p. 179]

$$\boldsymbol{\rho}_{\boldsymbol{u}} = (\boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A})_d \times [(\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})_d]^{-1}, \tag{2.12}$$

while in the case of one scale $u = \boldsymbol{a}'\boldsymbol{x}$ it becomes

$$\rho_{uu} = \frac{\text{var}(\boldsymbol{a}'\boldsymbol{B}\boldsymbol{\tau})}{\text{var}(\boldsymbol{a}'\boldsymbol{x})} = \frac{\boldsymbol{a}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}. \tag{2.13}$$

The formulas (2.12) and (2.13) can be given in other forms, where the matrix $\boldsymbol{\Psi}$ is explicitly present [17, p. 179]. It is often assumed that the measurement errors are uncorrelated, i.e., that $\boldsymbol{\Psi}$ is diagonal. We may use the notation $\boldsymbol{\Psi}_d$ to emphasize that $\boldsymbol{\Psi}$ is a diagonal matrix.

2.3 Other estimators

Other reliability estimators based on multidimensional measurement models have been suggested in the 1970s, but they have not replaced Cronbach's alpha. In the following, we briefly display two estimators, namely R_{cc} [11] and Ω [10], in order to show how they are related to Tarkkonen's rho.

Werts, Rock, Linn, and Jöreskog [11] consider the true score model $\boldsymbol{x} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}$, where $\boldsymbol{x}, \boldsymbol{\tau}$, and $\boldsymbol{\varepsilon}$ are random vectors of order p with the assumptions $\text{E}(\boldsymbol{\tau}) = \text{E}(\boldsymbol{x})$, $\text{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, and $\text{cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$. In addition, $\boldsymbol{\tau}$ is assumed to have an underlying factor model

$$\boldsymbol{\tau} = \boldsymbol{\Lambda}\boldsymbol{f} + \boldsymbol{\eta},$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times k}$ is the matrix of the factor loadings on k common factors $\boldsymbol{f}$, and $\boldsymbol{\eta}$ is the vector of the specific factors with $\text{E}(\boldsymbol{\eta}) = \boldsymbol{0}$. It is assumed that $\text{cov}(\boldsymbol{f}) = \boldsymbol{\Phi}$, $\text{cov}(\boldsymbol{\eta}) = \boldsymbol{\Theta}_d = \text{diag}(\theta_1^2, \ldots, \theta_p^2)$, $\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_d = \text{diag}(\psi_1^2, \ldots, \psi_p^2)$, and $\text{cov}(\boldsymbol{f}, \boldsymbol{\eta}) = \boldsymbol{0}$. Hence,

$$\text{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_d + \boldsymbol{\Psi}_d,$$

and the reliability estimator for the scale $\boldsymbol{a}'\boldsymbol{x}$ is [11, p. 934]

$$R_{cc} = \frac{\boldsymbol{a}'(\boldsymbol{\Sigma} - \boldsymbol{\Psi}_d)\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}} = \frac{\boldsymbol{a}'\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{\Theta}_d\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}.$$

Heise and Bohrnstedt [10] do no explicitly refer to the true score model and the factor model. However, their path-analytic approach [10, pp. 106–114] eventually leads to the context of the factor model, where the covariance matrix of the items is written, using the notation introduced above, in the form

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{U}_d,$$

where $\boldsymbol{U}_d = \boldsymbol{\Theta}_d + \boldsymbol{\Psi}_d$ represents the unique variance, the sum of the specific variance and the measurement error variance. The reliability estimator for the scale $\boldsymbol{a}'\boldsymbol{x}$ is [10, p. 115]

$$\Omega = \frac{\boldsymbol{a}'(\boldsymbol{\Sigma} - \boldsymbol{U}_d)\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}} = \frac{\boldsymbol{a}'\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}.$$

The main problem in R_{cc} and Ω is that in the factor model it is difficult to distinguish between the specific factors and the measurement errors [30, pp. 570–571]. If we assume that there is no item specific variance, i.e., assume that $\boldsymbol{\Theta}_d = \boldsymbol{0}$, then R_{cc} and Ω correspond to Tarkkonen's rho in the case of one scale. However, R_{cc} and Ω lack generality in respect of the definitions of the measurement model and the measurement scale.

3 Relations between Cronbach's alpha and Tarkkonen's rho

In this section we examine the relations between Cronbach's alpha and Tarkkonen's rho. We begin from the classical measurement scale, the unweighted sum, and then proceed to the weighted sum. We find it instructive to go through the unweighted sum first, although the results will follow immediately as special cases of the weighted sum. In each case, we prove that Tarkkonen's rho is the general method of estimating the reliability of a measurement scale, while Cronbach's alpha is its special case under certain, restricted conditions.

The results concerning Cronbach's alpha or KR-20 have been found in some form by numerous authors, see, e.g., [24,31,23,25] or [26,7,27,28,8,9]. Our aim is to collect these results together and present them with Tarkkonen's rho, using an up-to-date notation and style.

3.1 Case of the unweighted sum

In this subsection, we will refer to Cronbach's alpha and Tarkkonen's rho by

$$\alpha(\mathbf{1}) = \frac{p}{p-1}\left(1 - \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) \text{ and } \rho_{uu}(\mathbf{1}) = \frac{\mathrm{var}(\mathbf{1}'\boldsymbol{B\tau})}{\mathrm{var}(\mathbf{1}'\boldsymbol{x})} = \frac{\mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}.$$

We begin with the following theorem:

Theorem 1 *Let $\boldsymbol{V}$ be a symmetric nonnegative definite $p \times p$ matrix. Then*

$$\mathbf{1}'\boldsymbol{V}\mathbf{1} \geq \frac{p}{p-1}[\mathbf{1}'\boldsymbol{V}\mathbf{1} - \mathrm{tr}(\boldsymbol{V})], \tag{3.1}$$

i.e. (assuming $\mathbf{1}'\boldsymbol{V}\mathbf{1} \neq 0$),

$$1 \geq \frac{p}{p-1}\left(1 - \frac{\mathrm{tr}(\boldsymbol{V})}{\mathbf{1}'\boldsymbol{V}\mathbf{1}}\right).$$

The equality in (3.1) is obtained if and only if $\boldsymbol{V} = \delta^2\mathbf{1}\mathbf{1}'$ for some $\delta \in \mathbb{R}$.

PROOF. Let us rewrite (3.1) as

$$(p-1)\mathbf{1}'\boldsymbol{V}\mathbf{1} \geq p\mathbf{1}'\boldsymbol{V}\mathbf{1} - p\,\mathrm{tr}(\boldsymbol{V}),$$

that is,

$$\frac{\mathbf{1}'\boldsymbol{V}\mathbf{1}}{\mathbf{1}'\mathbf{1}} \leq \mathrm{tr}(\boldsymbol{V}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p, \tag{3.2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues of $\boldsymbol{V}$; we will denote $\lambda_i = \mathrm{ch}_i(\boldsymbol{V})$. Since

$$\max_{\boldsymbol{z}\neq\mathbf{0}} \frac{\boldsymbol{z}'\boldsymbol{V}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}} = \lambda_1 = \mathrm{ch}_1(\boldsymbol{V}),$$

the inequality (3.2) indeed holds. The equality in (3.2) means that

$$\lambda_1 \geq \frac{\mathbf{1}'\boldsymbol{V}\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \lambda_1 + \lambda_2 + \cdots + \lambda_p,$$

which holds if and only if $\lambda_2 = \cdots = \lambda_p = 0$ and vector $\mathbf{1}$ is the eigenvector of $\boldsymbol{V}$ with respect to λ_1, i.e., $\boldsymbol{V}$ is of the form $\boldsymbol{V} = \lambda_1\mathbf{1}\mathbf{1}'$. □

Puntanen and Styan [32, pp. 137–138] have proved Theorem 1 using orthogonal projectors. Another proof of Theorem 1 appears in [33, Lemma 4.1].

Theorem 1 implies immediately that

$$1 \geq \alpha(\mathbf{1}) = \frac{p}{p-1}\left(1 - \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) \tag{3.3}$$

for any $p \times p$ nonnegative definite matrix $\boldsymbol{\Sigma}$ (for which $\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} \neq 0$), and that the equality is obtained in (3.3) if and only if $\boldsymbol{\Sigma} = \delta^2\mathbf{1}\mathbf{1}'$ for some $\delta \in \mathbb{R}$.

Theorem 2 *Consider the measurement model* $\mathscr{M} = \{\boldsymbol{x}, \boldsymbol{B\tau}, \boldsymbol{B\Phi B}' + \boldsymbol{\Psi}_d\}$. *Then*

$$\rho_{uu}(\mathbf{1}) \geq \alpha(\mathbf{1}),$$

where the equality is obtained if and only if $\boldsymbol{B\Phi B}' = \delta^2\mathbf{1}\mathbf{1}'$ *for some* $\delta \in \mathbb{R}$, *i.e.,* $\boldsymbol{\Sigma} = \delta^2\mathbf{1}\mathbf{1}' + \boldsymbol{\Psi}_d$, *where* $\boldsymbol{\Psi}_d = \mathrm{diag}(\psi_1^2, \ldots, \psi_p^2)$.

PROOF. To prove that $\rho_{uu}(\mathbf{1}) \geq \alpha(\mathbf{1})$, we have to show that

$$\frac{\mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} \geq \frac{p}{p-1}\left(\frac{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} - \mathrm{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right). \tag{3.4}$$

Since $\boldsymbol{\Psi}_d$ is a diagonal matrix, we have

$$\begin{aligned}\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} - \mathrm{tr}(\boldsymbol{\Sigma}) &= \mathbf{1}'(\boldsymbol{B\Phi B}' + \boldsymbol{\Psi}_d)\mathbf{1} - \mathrm{tr}(\boldsymbol{B\Phi B}' + \boldsymbol{\Psi}_d) \\ &= \mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1} + \mathbf{1}'\boldsymbol{\Psi}_d\mathbf{1} - \mathrm{tr}(\boldsymbol{B\Phi B}') - \mathrm{tr}(\boldsymbol{\Psi}_d) \\ &= \mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1} - \mathrm{tr}(\boldsymbol{B\Phi B}'),\end{aligned}$$

and hence (3.4) is equivalent to

$$\mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1} \geq \frac{p}{p-1}[\mathbf{1}'\boldsymbol{B\Phi B}'\mathbf{1} - \mathrm{tr}(\boldsymbol{B\Phi B}')]. \tag{3.5}$$

In view of Theorem 1, (3.5) holds for every $\boldsymbol{B}$ (and every nonnegative definite $\boldsymbol{\Phi}$) and the equality is obtained if and only if $\boldsymbol{B\Phi B}' = \delta^2\mathbf{1}\mathbf{1}'$ for some $\delta \in \mathbb{R}$. $\square$

From Theorems 1 and 2 we can conclude that

$$\alpha(\mathbf{1}) \leq \rho_{uu}(\mathbf{1}) \leq 1.$$

The assumptions of Tarkkonen's rho ensure that $\rho_{uu}(\mathbf{1}) > 0$. However, $\alpha(\mathbf{1})$ may tend negative, because the original assumptions made in the derivation of KR-20 (2.4) and mostly inherited in $\alpha(\mathbf{1})$ are easily violated.

3.2 Case of the weighted sum

In this subsection, we will refer to Cronbach's alpha and Tarkkonen's rho by

$$\alpha(\boldsymbol{a}) = \frac{p}{p-1}\left(1 - \frac{\boldsymbol{a}'\boldsymbol{\Sigma}_d\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}\right) \text{ and } \rho_{uu}(\boldsymbol{a}) = \frac{\boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}.$$

Before proceeding onwards, we prove the following theorem:

Theorem 3 *Let $\boldsymbol{V}$ be a symmetric nonnegative definite $p \times p$ matrix with $\boldsymbol{V}_d = \mathrm{diag}(\boldsymbol{V})$ being positive definite. Then*

$$\max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{V}_d\boldsymbol{a}} = \mathrm{ch}_1(\boldsymbol{V}_d^{-1/2}\boldsymbol{V}\boldsymbol{V}_d^{-1/2}) = \mathrm{ch}_1(\boldsymbol{R}_V), \tag{3.6}$$

where $\boldsymbol{R}_V = \boldsymbol{V}_d^{-1/2}\boldsymbol{V}\boldsymbol{V}_d^{-1/2}$, i.e., $\boldsymbol{R}_V$ can be considered as a correlation matrix. Moreover,

$$\frac{\boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{V}_d\boldsymbol{a}} \leq p \quad \textit{for all } \boldsymbol{a} \in \mathbb{R}^p, \tag{3.7}$$

where the equality is obtained if and only if $\boldsymbol{V} = \delta^2\boldsymbol{q}\boldsymbol{q}'$ for some $\delta \in \mathbb{R}$ and some $\boldsymbol{q} = (q_1, \dots, q_p)'$, and $\boldsymbol{a}$ is a multiple of $\tilde{\boldsymbol{a}} = (1/q_1, \dots, 1/q_p)'$.

PROOF. We first note that

$$\begin{aligned}
\max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{V}_d\boldsymbol{a}} &= \max_{\boldsymbol{a}\neq\boldsymbol{0}} \frac{\boldsymbol{a}'\boldsymbol{V}_d^{1/2}\boldsymbol{V}_d^{-1/2}\boldsymbol{V}\boldsymbol{V}_d^{-1/2}\boldsymbol{V}_d^{1/2}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{V}_d^{1/2}\boldsymbol{V}_d^{1/2}\boldsymbol{a}} \\
&= \max_{\boldsymbol{z}\neq\boldsymbol{0}} \frac{\boldsymbol{z}'\boldsymbol{V}_d^{-1/2}\boldsymbol{V}\boldsymbol{V}_d^{-1/2}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}} \\
&= \mathrm{ch}_1(\boldsymbol{V}_d^{-1/2}\boldsymbol{V}\boldsymbol{V}_d^{-1/2}) \\
&= \mathrm{ch}_1(\boldsymbol{R}_V) := \mu_1.
\end{aligned}$$

It is obvious that the largest eigenvalue μ_1 of a $p \times p$ correlation matrix $\boldsymbol{R}_V$ is $\leq p$ and clearly $\mu_1 = p$ if and only if $\boldsymbol{R}_V = \boldsymbol{1}\boldsymbol{1}'$, i.e., $\boldsymbol{V}$ must be of the form $\boldsymbol{V} = \delta^2\boldsymbol{q}\boldsymbol{q}'$ for some $\delta \in \mathbb{R}$ and $\boldsymbol{q} = (q_1, \dots, q_p)' \in \mathbb{R}^p$. It is easy to conclude that if $\boldsymbol{V} = \delta^2\boldsymbol{q}\boldsymbol{q}'$, then the equality in (3.7) is obtained if and only if $\boldsymbol{a}$ is a multiple of $\tilde{\boldsymbol{a}} = \boldsymbol{V}_d^{-1/2}\boldsymbol{1} = \frac{1}{\delta}(1/q_1, \dots, 1/q_p)'$. □

Note that the maximum in (3.6) is obtained if and only if $\boldsymbol{a}$ is a multiple of $\tilde{\boldsymbol{a}} = \boldsymbol{V}_d^{-1/2}\boldsymbol{t}_1$, where $\boldsymbol{t}_1$ is the eigenvector of $\boldsymbol{R}_V$ with respect to the largest eigenvalue μ_1. In other words, $\tilde{\boldsymbol{a}}$ is a solution to the equation

$$\boldsymbol{V}\boldsymbol{a} = \mu_1\boldsymbol{V}_d\boldsymbol{a},$$

while $\boldsymbol{t}_1$ satisfies

$$\boldsymbol{R}_V\boldsymbol{t}_1 = \mu_1\boldsymbol{t}_1.$$

Clearly, when $\boldsymbol{R}_V = \boldsymbol{1}\boldsymbol{1}'$, the vector $\boldsymbol{t}_1$ is a multiple of $\boldsymbol{1}$.

Theorem 4 *Using the notation above,*

$$\alpha(\boldsymbol{a}) = \frac{p}{p-1}\left(1 - \frac{\boldsymbol{a}'\boldsymbol{\Sigma}_d\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}\right) \leq \frac{p}{p-1}\left(1 - \frac{1}{\mathrm{ch}_1(\boldsymbol{R}_\Sigma)}\right) \quad \textit{for all } \boldsymbol{a} \in \mathbb{R}^p.$$

Moreover,

$$\alpha(\boldsymbol{a}) \leq 1 \quad \textit{for all } \boldsymbol{a} \in \mathbb{R}^p. \tag{3.8}$$

The equality in (3.8) *is obtained if and only if* $\boldsymbol{\Sigma} = \delta^2 \boldsymbol{q}\boldsymbol{q}'$ *for some* $\delta \in \mathbb{R}$ *and some* $\boldsymbol{q} = (q_1, \ldots, q_p)'$*, and* $\boldsymbol{a}$ *is a multiple of* $\tilde{\boldsymbol{a}} = (1/q_1, \ldots, 1/q_p)'$.

PROOF. The proof comes at once from Theorem 3. □

Theorem 5 *Consider the measurement model* $\mathscr{M} = \{\boldsymbol{x}, \boldsymbol{B\tau}, \boldsymbol{B\Phi B}' + \boldsymbol{\Psi}_d\}$. *Then,*

$$\alpha(\boldsymbol{a}) \leq \rho_{uu}(\boldsymbol{a}) \textit{ for all } \boldsymbol{a} \in \mathbb{R}^p,$$

and the equality is obtained if and only if $\boldsymbol{B\Phi B}' = \delta^2 \boldsymbol{q}\boldsymbol{q}'$ *for some* $\delta \in \mathbb{R}$ *and some* $\boldsymbol{q} = (q_1, \ldots, q_p)'$*, i.e.,* $\boldsymbol{\Sigma} = \delta^2 \boldsymbol{q}\boldsymbol{q}' + \boldsymbol{\Psi}_d$*, and* $\boldsymbol{a}$ *is a multiple of* $\tilde{\boldsymbol{a}} = (1/q_1, \ldots, 1/q_p)'$.

PROOF. Our claim is

$$\frac{p}{p-1}\left(1 - \frac{\boldsymbol{a}'\boldsymbol{\Sigma}_d\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}\right) \leq \frac{\boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}},$$

i.e.,

$$\frac{p}{p-1}(\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} - \boldsymbol{a}'\boldsymbol{\Sigma}_d\boldsymbol{a}) \leq \boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a}. \tag{3.9}$$

Substituting $\boldsymbol{\Sigma} = \boldsymbol{B\Phi B}' + \boldsymbol{\Psi}_d$ (3.9) becomes

$$p[\boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{\Psi}_d\boldsymbol{a} - \boldsymbol{a}'(\boldsymbol{B\Phi B}')_d\boldsymbol{a} - \boldsymbol{a}'\boldsymbol{\Psi}_d\boldsymbol{a}] \leq (p-1)\boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a},$$

which simplifies into the form

$$\frac{\boldsymbol{a}'\boldsymbol{B\Phi B}'\boldsymbol{a}}{\boldsymbol{a}'(\boldsymbol{B\Phi B}')_d\boldsymbol{a}} \leq p.$$

The proof is now completed using Theorem 3. □

From Theorems 4 and 5 we can conclude that

$$\alpha(\boldsymbol{a}) \leq \rho_{uu}(\boldsymbol{a}) \leq 1 \text{ for all } \boldsymbol{a} \in \mathbb{R}^p.$$

The assumptions of Tarkkonen's rho ensure again that $\rho_{uu}(\boldsymbol{a}) > 0$, but similarly $\alpha(\boldsymbol{a})$ may tend negative. An additional reason is the introduction of the weights, since the basis of $\alpha(\boldsymbol{a})$, the original formula of KR-20 (2.4), was derived only for an unweighted sum.

4 Discussion and conclusions

Since the 1950s, Cronbach's alpha has become a routine, often referred to as "a popular method to measure reliability" [16], although there have been obvious problems caused by the strict assumptions inherited from its predecessors. In practical use of Cronbach's alpha, violation of the assumptions is unavoidable, and it leads to the well-known result of underestimation [7–9]. If reliability is underestimated, biased conclusions are made about the measurement, and any inference based on the reliability estimates, such as correction for attenuation [20], will also be biased or even useless [23, p. 276].

It is important for a measurement scale to have a high reliability, because the further analyses will then be based mainly on the true variation instead of random measurement errors. Of course, the question of validity should still be the primary concern. Unfortunately, the strict assumptions behind Cronbach's alpha have lead in the opposite direction: maximizing the reliability of the scale by discarding any "unsuitable" items. The applied criterion is the internal consistency, which requires all items to be equally good indicators of the trait under study. Some statistical program packages even support this procedure by reporting *alpha if item deleted* statistics. Combined with the underestimation problems of Cronbach's alpha, it is clear that this approach has unfavourable consequences.

In the 1930s, the fundamentals of the multiple factor analysis were established by Thurstone [34], and the researchers became more conscious of the multidimensionality of the psychological tests. It was noticed that the internal consistency, which was inherited from Spearman's one-factor theory [2,3] and implied by the method known as item analysis, was not a sufficient criterion for creating reliable tests [35–38]. Most empirical problems are multidimensional, and it is difficult to develop items that measure only one dimension. Indeed, the most striking problem of Cronbach's alpha is its built-in assumption of one-dimensionality. Using the notation established in the previous section, we can summarize three variants of the one-dimensional model that has dominated the test theory of psychometrics:

$$\begin{aligned}
&\mathscr{M}_1 = \{\boldsymbol{x}, \mathbf{1}\tau, \varphi^2\mathbf{1}\mathbf{1}' + \psi^2\boldsymbol{I}_p\}, \text{ where } \varphi^2 = \mathrm{var}(\tau) \text{ and } \psi^2 = \mathrm{var}(\varepsilon),\\
&\mathscr{M}_2 = \{\boldsymbol{x}, \mathbf{1}\tau, \varphi^2\mathbf{1}\mathbf{1}' + \boldsymbol{\Psi}_d\}, \text{ and}\\
&\mathscr{M}_3 = \{\boldsymbol{x}, \boldsymbol{b}\tau, \varphi^2\boldsymbol{b}\boldsymbol{b}' + \boldsymbol{\Psi}_d\}, \text{ where } \boldsymbol{b} \in \mathbb{R}^p.
\end{aligned}$$

In the literature, $\mathscr{M}_1$ is known as the parallel model [2], $\mathscr{M}_2$ is known as the τ-equivalent model [7], and $\mathscr{M}_3$ is known as the congeneric model [39]. Obviously, the models are special cases of $\mathscr{M} = \{\boldsymbol{x}, \boldsymbol{B}\boldsymbol{\tau}, \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{\Psi}\}$, the general, multidimensional measurement model [17].

According to Theorem 2, Cronbach's alpha is a valid estimator of reliability under $\mathscr{M}_1$ and $\mathscr{M}_2$, when the scale is the unweighted sum. In contrast, under $\mathscr{M}_3$ this holds only in the rare case where the scale weights are inverses of the model weights, as Theorem 5 shows. In practice, this would reduce $\mathscr{M}_3$ to $\mathscr{M}_2$ thus "standardizing" the effect of the true score. However, fine-tuning the assumptions of the models $\mathscr{M}_1$, $\mathscr{M}_2$, and $\mathscr{M}_3$ does not enhance the basic problem of one-dimensionality.

Generalizations of the one-dimensional model has been considered either by working with the factor model [10], or by combining it with the true score model [11]. The general measurement model $\mathscr{M}$ can also be seen as a combination of the true score model and the factor model. However, instead of the classical one-to-one correspondence between the observed items and the true scores, the common factors are directly associated with the true scores. The specific factors are interpreted either as measurement errors or as true scores, depending on the identification [17, pp. 176–177]. The measurement model is needed to specify the multidimensional structure of the measurement, but it must be stressed that the reliability is a property of the measurement scale. The reliability of the true scores (or the factors) is clearly undefined, although the phrase "reliability of the factor" seems to appear quite commonly in the literature.

In the measurement framework of Tarkkonen and Vehkalahti [17] the measurement scale is defined in broad terms, connecting the framework with the methods of multivariate statistical analysis [17, p. 173]. A general aim of these methods is to create various, possibly multidimensional scales, e.g., discriminant functions, canonical variates, or factor scores. The reliability of such scales can be estimated by Tarkkonen's rho without violating any assumptions. Our conclusion is that Tarkkonen's rho is a better alternative for Cronbach's alpha for the estimation of reliability.

Finally we note that throughout this paper we have been working with the population values, examining the methods of estimation of the "true reliability". Another question of estimation emerges when the population matrices are replaced by their sample estimators. Apparently, this will be a subject of a further study.

Acknowledgements

We are grateful to Jarkko M. Isotalo, Bikas K. Sinha, and Maria Valaste for helpful discussions.

References

[1] F. M. Lord, M. R. Novick, Statistical Theories of Mental Test Scores, Addison-Wesley, London, 1968.

[2] C. Spearman, The proof and measurement of association between two things, American Journal of Psychology 15 (1904) 72–101.

[3] C. Spearman, General intelligence objectively determined and measured, American Journal of Psychology 15 (1904) 201–293.

[4] W. A. Fuller, Measurement Error Models, Wiley, New York, 1987.

[5] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika 16 (1951) 297–334.

[6] L. J. Cronbach, W. Hartmann, A note on negative reliabilities, Educational and Psychological Measurement 14 (1954) 342–346.

[7] M. R. Novick, C. Lewis, Coefficient alpha and the reliability of composite measurements, Psychometrika 32 (1967) 1–13.

[8] D. J. Armor, Theta reliability and factor scaling, in: H. L. Costner (Ed.), Sociological Methodology, Jossey-Bass, San Francisco, 1974, pp. 17–50.

[9] V. L. Greene, E. G. Carmines, Assessing the reliability of linear composites, in: K. F. Schuessler (Ed.), Sociological Methodology, Jossey-Bass, San Francisco, 1979, pp. 17–50.

[10] D. R. Heise, G. W. Bohrnstedt, Validity, invalidity and reliability, in: E. F. Borgatta, G. W. Bohrnstedt (Eds.), Sociological Methodology, Jossey-Bass, San Francisco, 1970, pp. 104–129.

[11] C. E. Werts, R. D. Rock, R. L. Linn, K. G. Jöreskog, A general method of estimating the reliability of a composite, Educational and Psychological Measurement 38 (1978) 933–938.

[12] D. J. Weiss, M. L. Davison, Test theory and methods, Annual Review of Psychology 32 (1981) 629–658.

[13] R. R. Wilcox, Robust generalizations of classical test reliability and Cronbach's alpha, The British Journal of Mathematical and Statistical Psychology 45 (1992) 239–254.

[14] D. W. Zimmerman, B. D. Zumbo, C. Lalonde, Coefficient alpha as an estimate of test reliability under violation of two assumptions, Educational and Psychological Measurement 53 (1993) 33–49.

[15] K. Barchard, A. R. Hakstian, The effects of sampling model on inference with coefficient alpha, Educational and Psychological Measurement 57 (1997) 893–905.

[16] A. Christmann, S. V. Aelst, Robust estimation of Cronbach's alpha, Journal of Multivariate Analysis.

[17] L. Tarkkonen, K. Vehkalahti, Measurement errors in multivariate measurement scales, Journal of Multivariate Analysis 96 (2005) 172–189.

[18] S. F. Blinkhorn, Past imperfect, future conditional: Fifty years of test theory, The British Journal of Mathematical and Statistical Psychology 50 (1997) 175–185.

[19] G. F. Kuder, M. W. Richardson, The theory of the estimation of test reliability, Psychometrika 2 (1937) 151–160.

[20] C. Spearman, Correlation calculated from faulty data, British Journal of Psychology 3 (1910) 271–295.

[21] W. Brown, Some experimental results in the correlation of mental abilities, British Journal of Psychology 3 (1910) 296–322.

[22] L. J. Cronbach, On estimates of test reliability, The Journal of Educational Psychology 34 (1943) 485–494.

[23] L. Guttman, A basis for analyzing test-retest reliability, Psychometrika 10 (1945) 255–282.

[24] C. I. Mosier, On the reliability of a weighted composite, Psychometrika 8 (1943) 161–168.

[25] F. M. Lord, Some relations between Guttman's principal components of scale analysis and other psychometric theory, Psychometrika 23 (1958) 291–296.

[26] H. F. Kaiser, J. Caffrey, Alpha factor analysis, Psychometrika 30 (1965) 1–14.

[27] P. M. Bentler, Alpha-maximized factor analysis (alphamax): its relation to alpha and canonical factor analysis, Psychometrika 33 (1968) 335–345.

[28] R. P. McDonald, The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis, The British Journal of Mathematical and Statistical Psychology 23 (1970) 1–21.

[29] L. Tarkkonen, On Reliability of Composite Scales, no. 7 in Statistical Studies, Finnish Statistical Society, Helsinki, Finland, 1987.

[30] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd Edition, Wiley, New Jersey, 2003.

[31] R. J. Wherry, R. H. Gaylord, The concept of test and item reliability in relation to factor pattern, Psychometrika 8 (1943) 247–264.

[32] S. Puntanen, G. P. H. Styan, Matrix tricks for linear statistical models: our personal Top Thirteen, Research report A 345, Dept. of Mathematics, Statistics & Philosophy, University of Tampere, Tampere, Finland (2003).

[33] K. Vehkalahti, Reliability of Measurement Scales, no. 17 in Statistical Research Reports, Finnish Statistical Society, Helsinki, Finland, 2000.

[34] L. L. Thurstone, Multiple factor analysis, Psychological Review 38 (1931) 406–427.

[35] U. Handy, T. F. Lentz, Item value and test reliability, Journal of Educational Psychology 25 (1934) 703–708.

[36] J. Zubin, The method of internal consistency for selecting test items, Journal of Educational Psychology 25 (1934) 345–356.

[37] C. I. Mosier, A note on item analysis and the criterion of internal consistency, Psychometrika 1 (1936) 275–282.

[38] M. W. Richardson, Notes on the rationale of item analysis, Psychometrika 1 (1936) 69–76.

[39] K. G. Jöreskog, Statistical analysis of sets of congeneric tests, Psychometrika 36 (1971) 109–133.