The Importance of the Assumption of Uncorrelated Errors in Psychometric Theory

Educational and Psychological
Measurement

2015, Vol. 75(4) 634–647

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164414548217

epm.sagepub.com

Tenko Raykov¹, George A. Marcoulides², and
Thanos Patelis³

Abstract

A critical discussion of the assumption of uncorrelated errors in classical psychometric theory and its applications is provided. It is pointed out that this assumption is essential for a number of fundamental results and underlies the concept of parallel tests, the Spearman–Brown’s prophecy and the correction for attenuation formulas as well as the discrepancy between observed and true correlations, and the upper bound property of the reliability index with respect to validity. These relationships are shown not to hold if the errors of considered pairs of tests are correlated. The assumption of lack of error correlation is demonstrated not to be testable using standard covariance structure analysis for pairs of indivisible measures evaluating the same true score with identical error variances.

Keywords

attenuation, covariance structure analysis, error score, observed correlation, parallel tests, reliability, Spearman–Brown prophecy formula, true correlation, true score, uncorrelated errors, validity

¹Michigan State University, East Lansing, MI, USA

²University of California at Santa Barbara, Santa Barbara, CA, USA

³National Center for the Improvement of Educational Assessment, Dover, NH, USA

Corresponding Author:

Tenko Raykov, Measurement and Quantitative Methods, Michigan State University, 443A Erickson Hall,
East Lansing, MI 48824, USA.

Email: raykov@msu.edu

A critical revisit of classical psychometric theory reveals an important assumption that has received surprisingly little attention in the methodological and substantive literature across the educational, behavioral, social, marketing, business, and biomedical disciplines. This essential assumption, which underlies a number of its fundamental concepts and relationships, concerns the stipulation of uncorrelated errors associated with observed measures under consideration. Specifically, instrumental measurement concepts have been classically related directly or indirectly to the notion of parallel tests, yet these need to assume also that their error terms are uncorrelated in order to possess some of the properties for which they are widely referred to. Furthermore, the same assumption is essential for the validity of the popular Spearman–Brown prophecy formula. Moreover, the relationship between observed and true correlations for two given measures is grounded in the stipulation that their errors do not correlate, as is the popular correction for attenuation formula relating the two correlations. Last, the frequently cited feature of the reliability index as an upper bound of validity depends critically on the same presumption of uncorrelated errors for a test and criterion of interest.

Although this assumption of uncorrelated error terms in effect features through much of classical psychometric theory and its applications, we believe that it has not received the emphasis it deserves. The aim of the present note is to bridge this gap by highlighting the importance of this assumption and demonstrating its relevance for the validity and meaningfulness of the above concepts and relationships, as well as to show that they need not hold if it is violated. We also point out that it is not possible to test this assumption for pairs of indivisible measures using standard, currently routinely employed covariance structure analysis.

The plan of the article is as follows. The next section outlines the background, notation, and assumptions of the remainder of this discussion. The subsequent section shows the failure of major concepts and relationships in classical psychometric theory when the error terms of involved measures correlate. It is then demonstrated that for two single-component measures with equal loadings on a common construct and equal error variances, it is not possible to test with empirical data whether their errors correlate using the conventional covariance structure analysis approach. The conclusion section summarizes the findings of the article and makes some recommendations.

Background, Notation, and Assumptions

This note instrumentally uses the classical test theory (CTT) framework, which is most rigorously presented in Zimmerman (1975). The reason for adopting this framework is the fact that the notion of parallel tests, which is of direct or indirect importance for some of the concepts and relationships of concern in the article, was developed within CTT in the past century (e.g., Lord & Novick, 1968). Similarly, the remaining notions and relations of interest below were advanced a number of decades ago within the same framework and obtained quickly prominence in the

methodological and substantive literature. We observe also that in the setting of relevance in the article, the CTT-based models considered are not empirically distinguishable from corresponding single-factor models (with appropriate restrictions) using standard covariance structure analysis, as they imply the same covariance and mean structures (e.g., Raykov & Marcoulides, 2006).

Throughout the present article, we consider a pair of measures, denoted X_1 and X_2 (and frequently also referred to as “tests”), and assume that they are indivisible, that is, not divisible further into components. That is, X_1 and X_2 consist each of single components rather than represent sums (weighted or unweighted) of scores on more elementary components that make up either of them. Designating by T_1 and T_2 the true scores associated with the observed scores X_1 and X_2 , respectively, and by E_1 and E_2 their corresponding error scores, according to CTT the well-known decomposition

$$X_i = T_i + E_i \quad (1)$$

holds ($i = 1, 2$).

For Equation (1) to be applicable, as discussed in detail in Zimmerman (1975), one only needs to assume the existence of the mean of the pertinent propensity distribution associated with each measure (that is typically considered a random variable). For the existence of the mean of each of X_1 and X_2 , it is sufficient that their variances exist, which may be considered a rather mild if at all restrictive condition in practice (e.g., Lord & Novick, 1968). We emphasize that within CTT there is no assumption, or any need for such, that the errors E_1 and E_2 correlate, or that they alternatively do not correlate.¹ Whether these errors correlate or not, is itself an assumption that needs to be adopted in any particular utilization of CTT in an empirical setting or a model based on it (e.g., Zimmerman, 1975; see also Raykov & Marcoulides, 2011). Furthermore, the true and error scores associated with a given observed score, X say, do not correlate by the definition of true score, that is, $\text{Corr}(T, E) = 0$ holds for any observed and corresponding true score, as follows from the construction of the true score T (and hence of the error score E) from its observed score X , where $\text{Corr}(\cdot, \cdot)$ denotes correlation. Also, for the purposes of this note, we assume in the remainder that true and error variances are positive, that is, there are individual differences in a studied population on each true and error score associated with any observed measure in question, and hence also on the latter measure. In most contemporary empirical behavioral and social research, this is a stipulation that can be considered practically fulfilled as well.

To summarize the preceding discussion, within the framework of CTT there is no assumption of uncorrelated error terms pertaining to two distinct observed scores. Any particular utilization of CTT with two or more manifest measures, however, whether in theoretical discourse or in an empirical application, must assume either that their pertinent error terms are uncorrelated, or that they are correlated, since CTT per se does not make (or need to make) either of these two assumptions. Whether a user of CTT makes the assumption then of uncorrelated errors or alternatively of

correlated errors associated with at least two observed scores, is an issue that he or she needs to resolve and explicate.

In the next section, we attend to several major results in classical psychometric theory, which no longer hold if the assumption of uncorrelated errors is violated.

The Essential Assumption of Uncorrelated Errors

Parallel Tests and Reliability

Consider X_1 and X_2 as two given tests and assume we were interested in examining whether they were parallel. A fact that in our opinion has not received sufficient attention in the literature, is that their error scores must be uncorrelated if the tests are to be useful in most classical psychometric applications. Specifically, for X_1 and X_2 to be parallel—with all implications typically drawn from that within the framework of classical psychometrics and its utilizations—it must be true that for their errors $\text{Corr}(E_1, E_2) = 0$ holds. Otherwise, as can be readily seen below, the correlation of the two tests cannot equal the reliability coefficient of either of them, yet this identity is used instrumentally in many discussions in theoretical and empirical research. (We refer to the reliability coefficient usually as “reliability” in the rest of this article.)

Indeed, suppose we define a pair of measures X_1 and X_2 as parallel only based on the following two requirements (see also Equation 1):

$$T_1 = T_2 = T \quad \text{and} \quad \text{Var}(E_1) = \text{Var}(E_2) = \theta \quad (2)$$

say, where $\text{Var}(\cdot)$ denotes variance and T and θ designate the common true score and common error variance, respectively. That is, Equations (2) state that the two tests measure the same true score with the same imprecision (as captured by their error variances).

Equations (2), which are oftentimes used to define parallel measures in classical psychometric treatments, entail that these measures are also congeneric (e.g., Jöreskog, 1971), that is, the equations

$$X_1 = T + E_1 \quad \text{and} \quad X_2 = T + E_2 \quad (3)$$

hold, or in other words the tests have each a loading of 1 on their common true score. (We also note in passing that from the first of Equations 2, or from Equations 3, the equality of the means of X_1 and X_2 also follows, due to that of the error terms being 0; see also Zimmerman, 1975.) From Equations (3), it is seen that then parallel tests are a special case of congeneric measures. (As is well known, the latter are defined as measures of the same true score with possibly different units of measurement and error variances; e.g., Jöreskog, 1971.) Similarly, and along the lines of the discussion in the preceding section, it does not follow from the definition also of congeneric measures that their error scores are uncorrelated. Rather, one needs to assume their error uncorrelatedness, or alternatively their error correlation (for

particular pairs of them), in any application of a set of congeneric measures (e.g., Raykov & Marcoulides, 2011).

While the assumptions in Equations (2) imply readily that the observed variances of the two measures are the same, as are their true variances and reliability coefficients (defined traditionally as the ratios of true to observed measure variance), their correlation no longer equals the reliability of either test as soon as their error terms correlate. This is because for the measure covariance obviously

$$\text{Cov}(X_1, X_2) = \text{Var}(T) + \text{Cov}(E_1, E_2) = \text{Var}(T) + \psi \neq \text{Var}(T)$$

is true then, where $\text{Cov}(\cdot, \cdot)$ denotes covariance and ψ designates that of the error terms in question (with $\psi \neq 0$ in the current discussion). Hence, denoting by $\rho(\cdot)$ reliability, evidently

$$\text{Corr}(X_1, X_2) > \rho(X_1) = \rho(X_2) = \rho \quad (4)$$

holds whenever $\psi > 0$, with ρ symbolizing the common reliability coefficient. Conversely,

$$\text{Corr}(X_1, X_2) < \rho \quad (5)$$

is obviously true in case $\psi < 0$. (Both inequalities 4 and 5 hold also due to the measure variances being equal, as mentioned above.) That is, when their error terms correlate, the correlation of two measures fulfilling just the two assumptions in Equations (2) no longer equals the reliability of each one of them, and hence their correlation cannot be used as a credible index informing about the reliability of either measure.

This discussion shows that unless one adds the requirement

$$\text{Corr}(E_1, E_2) = 0, \quad (6)$$

two tests that only satisfy Equations (2) should not be really referred to as parallel. This is because they do not fulfill then an essential property for which the concept of parallel tests was advanced in the first place, namely, that their correlation is equal to the reliability of either of them. In other words, the assumption (6) of uncorrelated errors is an essential stipulation for two measures—in addition to them evaluating the same true score with the same imprecision (Equations 2)—to be indeed parallel as intended by classical psychometrics (see also the Conclusion section).

Based on the above developments, we believe that a complete definition of parallel tests is as such measures of the same true score with identical error variances, whose error terms are uncorrelated. We thus propose to use this complete definition of parallel tests whenever using them in theoretical or empirical discussions. This is the definition of parallel tests that also underlies the present article and is essential for the typical applications of parallel tests in educational, behavioral, and social research (see also the Conclusion section).

The Spearman–Brown Prophecy Formula

When the error terms associated with two tests correlate, the Spearman–Brown “prophesy” (SBP) formula does not hold even if these measures satisfy Equations (2), as we show next. Indeed, denoting then by ρ_Z the reliability of the sum score $Z = X_1 + X_2$, from Equations (3) it readily follows (e.g., Bollen, 1989) that

$$\rho_Z = \frac{2\text{Var}(T)}{2\text{Var}(T) + \theta + \psi}. \quad (7)$$

Designate now by $\rho_{Z,SB}$ the reliability of the sum score Z according to the SBP formula. As is well known, the latter states (cf., e.g., McDonald, 1999)

$$\rho_{Z,SB} = 2\rho/(1 + \rho). \quad (8)$$

Given that $\rho = \text{Var}(T)/(\text{Var}(T) + \theta)$ here, after some straightforward algebra based on the congeneric model (3), the equality

$$\rho_{Z,SB} = \frac{2\text{Var}(T)}{2\text{Var}(T) + \theta} \quad (9)$$

follows from (8), whose right-hand side contradicts that of (7), however (recall that $\psi \neq 0$ is assumed throughout this section).

Hence, whenever the errors of two observed measures satisfying (only) Equations (2) correlate positively, the SBP formula will actually over-predict the reliability of their sum (compare Equations 8 and 9); conversely, the SBP formula will underpredict the reliability of that sum whenever these errors correlate negatively. Thereby, the extent of this over- or underprediction depends on the magnitude of the error covariance relative to the common true variance: the larger the error covariance relative to the latter, the stronger this biasing feature of the SBP formula in case of correlated errors.

This discussion demonstrates that the assumption of uncorrelated errors is also essential for the validity of the popular SBP formula. Specifically, the SBP formula is no more applicable with correlated error scores of two tests that evaluate the same true score and share the same error variance.² We find that the last statement is also consistent with a recent criticism by Charter (2001) of the overly frequent reference to and use of the SBP formula in applied as well as theoretical measurement discussions in the educational, behavioral, and social disciplines.

Observed Versus True Correlation

The lower bound property of observed correlation with respect to true correlation for two given measures in the general case - that is, whether or not they are parallel, essentially tau-equivalent or even congeneric (Lord & Novick, 1968) - is widely referred to in discussions and applications of classical psychometrics. Far less

attention, however, has been given then to an essential assumption for this property to hold—the lack of correlation for the error terms associated with these measures.

Indeed, if these errors do not correlate, then the aforementioned lower bound property holds, that is, the correlation of the two observed measures cannot exceed the correlation of their true scores (cf., e.g., Crocker & Algina, 2006). However, when the error terms of the two measures correlate, then their observed correlation need no longer be a lower bound of the correlation of their true scores, as can be readily seen next. To be concrete, if

$$X_1 = T_1 + E_1 \quad \text{and} \quad X_2 = T_2 + E_2 \quad (10)$$

are their corresponding CTT decompositions, with their error terms E_1 and E_2 being correlated, then

$$\begin{aligned} \text{Corr}(X_1, X_2) &= \text{Corr}(T_1 + E_1, T_2 + E_2) \\ &= \frac{\text{Cov}(T_1, T_2)}{[SD(X_1)SD(X_2)]} + \frac{\text{Cov}(E_1, E_2)}{[SD(X_1)SD(X_2)]}, \end{aligned} \quad (11)$$

where $SD(\cdot)$ denotes standard deviation. (By the definition of true score, for any measure its true score is uncorrelated with the error score of any other measure; e.g., Zimmerman, 1975.)

From Equation (11) it is readily seen that if $\text{Cov}(E_1, E_2) > 0$, then it can happen that observed correlation is in fact larger than true correlation, rather than the former being a lower bound of the latter. A simple example is readily observed when true and error variances are all equal to 1, while $\text{Cov}(E_1, E_2) = 0.75$ and $\text{Cov}(T_1, T_2) = 0.5$ say; then straightforward calculations show that

$$\text{Corr}(T_1, T_2) = 0.25 < \text{Corr}(X_1, X_2) = 0.62. \quad (12)$$

Inequality (12) is incompatible, however, with the frequently cited lower bound property of observed correlation with respect to true correlation—in fact, (12) is exactly the opposite of that property. (Note that in this example, all three involved covariance matrices—of the two true scores, of the two observed scores, and of the two error scores—are positive definite; e.g., Raykov & Marcoulides, 2008.)

To summarize, the assumption of error uncorrelatedness of two observed measures is also essential for the widely referred to property of observed correlation being a lower bound of true correlation, that is, of the empirical correlation of these measures being a lower bound of the correlation of their associated true scores.

The Correction-for-Attenuation Formula

The discussion in the last subsection is also closely related to the popular correction-for-attenuation (CFA) formula. Moreover, as shown next, the CFA formula (cf.

Crocker & Algina, 2006) need not be correct or applicable unless the error terms of two considered observed measures are uncorrelated. Indeed, if their error scores are allowed to be correlated, then from Equation (11) one sees that

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(T_1, T_2)}{[SD(X_1)SD(X_2)]} + \delta, \quad (13)$$

where $\delta = \text{Cov}(E_1, E_2)/[SD(X_1)SD(X_2)] \neq 0$.

For the first term in the right-hand side of Equation (13), obviously the following equality holds

$$\begin{aligned} & \frac{\text{Cov}(T_1, T_2)}{[SD(X_1)SD(X_2)]} \\ &= \text{Corr}(T_1, T_2) \sqrt{(\rho_1 \rho_2)}, \end{aligned} \quad (14)$$

where ρ_1 and ρ_2 denote the reliability coefficients of X_1 and X_2 , respectively (and “ $\sqrt{\cdot}$ ” designates positive square root of the following product in parentheses). We observe that the right-hand side of Equation (14) is precisely the right-hand side of the well-known CFA in case of uncorrelated errors (e.g., Allen & Yen, 2001). As discussed in classical psychometrics treatments, the CFA can be used in that case to “correct” an observed correlation in an attempt to obtain an estimate of the associated true score correlation that is typically of actual interest—from Equation (14), the true correlation is then the ratio of observed correlation to the product of the reliability indexes of the two measures involved. (Note that under the assumption of positive true and error variances made at the outset of this note, the reliability of either test cannot be 0, and hence the resulting ratio is well defined.)

Substituting the right-hand of Equation (14) into the right-hand side of (13), we see that

$$\text{Corr}(X_1, X_2) = \text{Corr}(T_1, T_2) \sqrt{(\rho_1 \rho_2)} + \delta. \quad (15)$$

Therefore, the popular CFA formula does not hold in case of correlated errors of the tests X_1 and X_2 , as it actually under-/overcorrects when errors are negatively/positively correlated. Thereby, the extent to which the CFA is violated in this sense and would be misleading if nonetheless used, is determined by the degree to which the error terms E_1 and E_2 correlate and the extent to which the observed measures capture individual differences (test variances).

By way of summary, the error uncorrelatedness assumption for two observed measures is similarly needed to hold in order for the popular CFA formula to be valid.

Reliability (Index) as a Lower Bound of Validity?

When the errors of a test under consideration and a criterion measure correlate, the reliability index also need not be an upper bound of the criterion validity coefficient.

The latter property, of validity being bounded above by the reliability index, has been frequently cited in classical psychometrics discussions and applications as a means of gauging the magnitude of criterion validity in validation studies (cf., e.g., Crocker & Algina, 2006). This property does hold when the error terms do not correlate, but need not hold and can be misleading if used whenever they correlate, as can be readily seen as follows.

To this end, we take in the preceding subsection X_1 as the test of interest and X_2 as the criterion of relevance, and note that we can employ Equation (15) next. To commence, we point out that with no loss of generality we can assume these two tests as congeneric because we are concerned in this subsection with showing an instance of a failure of the aforementioned, widely cited upper bound property of the reliability index with respect to criterion validity. Furthermore, for simplicity we can similarly assume that both tests have the same reliability, denoted ρ , and unitary variance. Then Equation (15) implies

$$\text{Corr}(X_1, X_2) = \rho + \text{Cov}(E_1, E_2). \quad (16)$$

With Equation (16) in mind, suppose we chose a reliability of say .81 for both the test and criterion in question. Then we easily notice from the right-hand side of (16) that $\text{Cov}(E_1, E_2)$ could be chosen to be at least 0.10 (and still with the property that the resulting error correlation is within the admissible range of -1 through 1), such that the right-hand side of Equation (16) becomes larger than the pertinent reliability index (which is 0.9 here). This choice will, however, invalidate the widely cited upper bound property of the reliability index with respect to criterion validity. That violation occurs because the latter validity coefficient is the left-hand side of (16) then, which as just shown is larger than the square root here of the first term on its right-hand side, that is, the reliability index of the test under consideration.

Therefore, when the error terms of a test and a criterion measure of interest correlate, the reliability index of the former need not be an upper bound of the criterion validity coefficient, contrary to the frequently referred to property of the reliability index being an upper bound of validity in classical psychometrics and its applications.

To summarize the developments so far in this article, the assumption of uncorrelated errors is essential for a proper definition of parallel tests, the SBP formula, the lower bound property of observed correlation with respect to true correlation, and for the CFA formula. With this special relevance of that error uncorrelateness assumption in mind, we address in the following section the query of whether one can test that assumption empirically, specifically in a setting of particular interest in classical psychometrics that is attended to next.

Can We Test for Lack of Error Correlation in Two “Parallel” Measures?

The preceding discussion of the essential assumption of vanishing error correlation raises the question of how to go about ascertaining that in a given empirical setting

one is not dealing with a case when a number of important concepts and relationships in classical psychometrics can fail. Unfortunately, as demonstrated next, it is not possible with standard and currently routinely used covariance structure analysis to test whether the errors correlate for two indivisible measures with the same true score and identical error variances.

To see this, we assume first that for two indivisible measures, X_1 and X_2 , Equations (2) hold with the added assumption of their uncorrelated errors. We denote by M1 the resulting model (that we referred to earlier as the “parallel test model”). We stress that M1 is a model based on the assumption of uncorrelated errors in addition to the stipulation of the two measures evaluating the same true score with the same imprecision (error variance); that is, the two measures are congeneric, with the same loadings on the common true score and identical error variances, as well as uncorrelated error scores. Then the implied covariance matrix by model M1 is

$$\Sigma_{(1)} = \begin{bmatrix} \phi + \theta & \phi \\ \phi & \phi + \theta \end{bmatrix}, \quad (17)$$

where ϕ symbolizes the common true score variance (e.g., Raykov & Marcoulides, 2006). We note that Equations (2) do not imply any restrictive parameterization of the mean structure of the two tests, irrespective of whether we add the lack of error covariance assumption or not, and therefore we need not be concerned with their mean structure in the rest of the present section.

If we relax now the uncorrelated error assumption, while still keeping the two assumptions in Equations (2), we obtain another model that we denote M2. We observe that M2 has one more parameter than M1. On denoting by ψ that error covariance, the implied covariance matrix by model M2 is as follows (we add an asterisk to the true and error variance symbols next, while using otherwise the same basic notation, in order to emphasize the difference in the two models):

$$\Sigma_{(2)} = \begin{bmatrix} \phi^* + \theta^* & \phi^* + \psi \\ \phi^* + \psi & \phi^* + \theta^* \end{bmatrix} \quad (18)$$

When comparing the implied covariance matrices by models M1 and M2, that is, the right-hand sides of Equations (17) and (18), it is readily observed that the two models reproduce the same covariance matrix with appropriate choices of their parameters (cf. Raykov & Penev, 1999). In particular, for any $\phi > 0$ and $\theta > 0$, such that (17) is positive definite, we can obviously choose (a) $\phi^* < \phi$ while $\phi^* > 0$, (b) $\theta^* = \phi + \theta - \phi^*$, and (c) $\psi = \phi - \phi^*$, such that the right-hand side of (18) yields the right-hand side of (17), that is, $\Sigma_{(2)} = \Sigma_{(1)}$ holds. Conversely, for any triple of numbers ϕ^* , θ^* , and ψ (with the first two being positive) such that $\Sigma_{(2)}$ is positive definite (which implies $\psi < \theta^*$), by setting $\phi = \phi^* + \psi$ and $\theta = \theta^* - \psi$, it follows that the right-hand side of (17) yields the right-hand side of (18), that is, $\Sigma_{(1)} = \Sigma_{(2)}$ holds.

Hence, the set of covariance matrices implied by Model M1 is identical to the set of covariance matrices implied by model M2. (We stress that the two models are not

equivalent since they have different numbers of parameters, as indicated above; e.g., Raykov & Penev, 1999.) Thus, a positive definite 2×2 covariance matrix may have been generated by a model satisfying Equations (2) with uncorrelated errors, but we have no way of differentiating that matrix from a positive definite 2×2 covariance matrix that has been generated by a model satisfying (2) yet with correlated errors for the two measures involved and appropriately chosen values of its parameters; and vice versa.

Therefore, for any pair of observed (indivisible) measures, the covariance matrix implied by Model M2 contains no information over and above that contained in the covariance matrix implied by Model M1, and conversely, while the only difference across the two models is the assumption of uncorrelated errors in M1. As a consequence, for any two indivisible observed measures fulfilling Equations (2), the assumption of uncorrelated errors is actually not testable using standard covariance structure analysis, owing to the fact that the latter approach cannot differentiate between Models M1 and M2.³

Conclusion

This article aimed to demonstrate that for a number of fundamental concepts, relationships, and results in classical psychometric theory to hold for a pair of observed measures, it is essential to make the assumption of their error terms being uncorrelated. When this assumption is not advanced or does not hold, that is, is violated, two measures that evaluate the same true score with the same imprecision (error variance) cannot and should not be really considered or referred to as parallel, since their correlation no longer equals the reliability of either of them. Similarly, the popular SBP formula cannot be relied on in such cases (see also Note 1). Moreover, one cannot trust the correction of attenuation formula then, the lower bound property of observed correlation with respect to true correlation, and the upper bound property of the reliability index with regard to criterion validity.

Unfortunately, this assumption of uncorrelated errors, while essential for the above concepts and relationships, cannot be tested for two indivisible manifest measures using standard covariance structure analysis (see also Raykov, Patelis, & Marcoulides, 2011, for a discussion of related limitations when examining other aspects of the question whether two given measures are parallel, including the case of them representing multiple-component measuring instruments). We therefore encourage future research into possible alternative analytic and/or modeling approaches to empirical examination of this critical assumption for a pair of indivisible (i.e., single-component) tests.⁴

With regard to the widely used concept of parallel tests in classical psychometrics and related treatments and applications, this article provides the grounds for the recommendation to routinely emphasize the assumption of uncorrelated error terms when discussing pairs of observed measures because this assumption is essential then, as elaborated in this note. This should, in particular, be done when referring to

the possibly most popular feature of parallel tests—namely, being associated with correlation that equals the reliability of each one of them. Based on the preceding developments in this note, we thus do not find it meaningful to involve the concept of parallel tests in theoretical or empirical discussions and research with pairs of measures without invoking this assumption of uncorrelated errors, especially when reliability of the individual measures or their sum is a focus of concern. In other words, two parallels should not be defined, or considered, as parallel by requiring only Equations (2) to hold, but rather by requiring the two equations in (2) to hold *and* additionally assuming uncorrelated errors.

Last but not least, this article also contributes to the following suggested recommendation: One need not attach to the concept of parallel tests the importance it receives and has received in the past in discussions of classical psychometrics and its applications. In fact, we submit that classical psychometric theory will not lose much of its relevance for the educational, behavioral, and social disciplines if the concept of parallel tests was left unused in it and its applications in empirical research (cf., e.g., Raykov & Marcoulides, 2011), at least until a routinely and widely applicable means was found for testing for uncorrelated errors in pairs of indivisible measures with identical true scores and error variances (see also Charter, 2001).

Acknowledgments

We thank R. Steyer for valuable discussions on classical test theory. We are grateful to an anonymous referee for critical comments on an earlier version of the article, which contributed considerably to its improvement.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. As discussed in the literature (e.g., Crocker & Algina, 2006; Williams & Zimmerman, 1996), errors of two congeneric tests may be correlated if the tests are presented ‘closely’ in time or space, for instance as items pertaining to the same paragraph in a reading test or referring to the same figure in a figural relations test or a geometry test, or administered under special conditions.
2. By analogy, one can show that the lack of error correlation is essential for the SBP formula also when considering more than 2 observed measures that evaluate the same true score with the same imprecision (error variance). Specifically, in general this formula will under- or over-predict the reliability of the sum of these measures depending on the relative

magnitude of the non-zero error covariances and their relationships to other parameters associated with the measures. (Special cases are then also possible, obviously, when positive error covariances cancel out with negative error covariances in the denominator of the sum score reliability coefficient and the SBP formula may happen to hold.)

3. Another way of realizing that the uncorrelated error assumption is untestable via standard covariance structure analysis, is as follows. Model M2 is not identified, which is readily shown along the same lines as in the present section, and it is the error covariance ψ that is an unidentified parameter in M2. Hence, infinitely many values of this covariance – in fact also including the value of 0 as shown by the preceding discussion in this section – are associated with the same (overall) fit to any positive definite covariance matrix of size 2×2 . This implies that a given 2×2 empirical covariance matrix cannot contain information that could be used to differentiate between all these possible values of the error covariance, leading to the lack of testability of the pertinent assumption of uncorrelated errors. Based on the developments in this section, we suggest that the error uncorrelatedness assumption is not testable also if one were to use for this purpose individual case residuals (cf. Raykov & Penev, 2014).
4. As indicated at the outset, this note is not concerned with settings characterized by (a) more than two observed measures or (b) a pair of measuring instruments (scales, tests) consisting of multiple components (for a discussion related to case (b), see Raykov et al., 2011).

References

- Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Charter, R. A. (2001). It is time to bury the Spearman-Brown “prophecy” formula for some common applications. *Educational and Psychological Measurement*, 61, 690-696.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. New York, NY: Taylor & Francis.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.
- Raykov, T., Patelis, T., & Marcoulides, G. A. (2011). Examining parallelism of sets of psychometric measures using latent variable modeling. *Educational and Psychological Measurement*, 71, 1047-1061.
- Raykov, T., & Penev, S. (1999). On structural equation model equivalence. *Multivariate Behavioral Research*, 34, 199-244.
- Raykov, T., & Penev, S. (2014). Latent growth curve models selection: The potential of individual case residuals. *Structural Equation Modeling*, 21, 20-30.

- Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, 20, 59-69.
- Zimmerman, D. W. (1975). Probability measures, Hilbert spaces, and the axioms of classical test theory. *Psychometrika*, 30, 221-232.