The Effects of Sample Size, Correlation Technique, and Factor Extraction Method on Reliability Coefficients

Örneklem Büyüklüğü, Korelasyon Tekniği ve Faktör Çıkarma Yönteminin Güvenirlik Katsayılarına Etkisi

Nuri DOĞAN[a], Abdullah Faruk KILIÇ[a]
[a]Hacettepe University Faculty of Education, Department of Educational Sciences Division of Educational Measurement and Evaluation, Ankara, Turkey

Öz

Bu araştırmanın amacı, güvenirlik katsayılarını örneklem büyüklüğüne (250, 500, 1000, 2500, 5000 ve 9773), EFA faktör çıkarma yöntemine (PCA, PA, ULS, WLS ve MLE); CFA kestirim yöntemine (UL, ML ve GL) ve korelasyon matrisine (Pearson, phi ve tetrakorik) göre karşılaştırmaktır. Bu amaçla temel araştırma yöntemine başvurulmuştur. Araştırma gerçek verilerle yürütülmüş olup araştırma verileri, 2014 yılında uygulanan Temel Eğitimden Ortaöğretime Geçiş sınavları Türkçe alt testine cevap veren öğrenci cevaplarından oluşturulmuştur. Araştırmada McDonald ω, McDonald ω_h, Maksimal güvenirlik, Armor Ѳ, Heise ve Bohrnstedt Ω, Revelle β ve Standartlaştırılmış Alfa katsayıları karşılaştırılmıştır. Araştırma sonucunda aynı korelasyon matrisi için örneklem büyüklüğünün çok değişikliğe neden olmadığı gözlenmiştir. Tetrakorik korelasyonla hesaplanan McDonald ω_h ve Revelle β katsayılarının 1'den büyük olduğu bazı koşulların bulunduğu araştırmanın bir diğer bulgusudur. Araştırma sonuçlarına göre bu katsayıların konjerik tek faktörlü yapılar için phi korelasyonuyla hesaplanmasını önerilmiştir. Diğer bulgular literatür eşliğinde tartışılmış ve önerilerde bulunulmuştur.

Abstract

This study aims to compare reliability coefficients according to sample size (250, 500, 1,000, 2,500, 5,000, and 9,773), EFA factor extraction methods (PCA, PA, ULS, WLS, and MLE), CFA estimation methods (UL, ML, and GL), and correlation matrices (Pearson, phi, and tetrachoric). Therefore, it employs a basic research method. The study was conducted with real data, and the data were collected from students' answers to a Turkish sub-test in the Test for Transition from Basic Education into Secondary Education administered in 2014. Within the scope of the study, McDonald ω, McDonald ω_h, maximal reliability, Armor Ѳ, Heise and Bohrnstedt Ω, Revelle β, and standardized alpha coefficients were compared. Consequently, it was found that sample size in the same correlation matrices did not lead to serious changes. It was also found that McDonald ω_h and Revelle β coefficients calculated with a tetrachoric correlation were bigger than 1 in some conditions. It was recommended in consequence that those coefficients should be calculated through phi correlations for congeric one-factor structures. Other findings obtained support the literature, and necessary suggestions are made.

1. Introduction

Scientists working in the field of social sciences often measure individuals' properties that cannot be directly observed. Those properties can be the ability to play basketball, attitudes towards a course, interest in professions, or any other cognitive variables, as well as mathematical achievement. Those directly unobservable properties are called "constructs" or "latent traits." In measuring those directly unobservable traits, behaviors thought to be associated with them are measured. Total scores obtained by assigning 1 to correct answers and 0 to incorrect answers in a mathematics achievement test could be an example for measuring directly unobservable latent traits through observable behaviors. However, it is not adequate simply to observe individuals' behaviors. The behaviors should be associated with latent traits, and individuals' scores should be consistent. The consistency of individuals' scores is related to reliability, and the relevance between behaviors and latent variable is related to validity (Meyer, 2010).

Considering reliability within the framework of classical test theory (CTT) and representing the score an individual i can receive from an item j as V_{ij}, then V_{ij} is composed of two components:

$$V_{ij} = t_{ij} + e_{ij} \qquad (1.1)$$

Where t_{ij} stands for an individual's real score, and e_{ij} stands for the error score. It is assumed that the average error score for an item in a sample is 0 and that the error scores are not related to the real score (Green & Yang, 2009). Besides, it can be said that individuals' observed scores slightly deviate from the actual scores according to this equation (Gulliksen, 1950). As the amount of deviation in individuals' observed scores from actual scores increases, reliability decreases; but as the amount of deviation decreases, reliability increases. Accordingly, reliability is described as

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} = \frac{\sigma_T^2}{\sigma_X^2} \qquad (1.2)$$

According to this, it may be said that reliability will increase as the observed score comes closer to the real score — that is to say, as the error decreases. The sources of error can be examined and the necessary precautions can be taken so as to reduce errors. As the sources of error change, the definitions of reliability can also change.

Reliability is defined in various ways. To exemplify a few for instance, if the scores taken from a test have high correlations with the actual scores, then the test is said to be reliable (Allen & Yen, 1979). Besides, reliability can also be considered as obtaining the same results when a test is administered again in the same conditions (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; Crocker & Algina, 2008).

There are a great number of reliability coefficients that have been developed on the basis of CTT and of which ways of calculation are different. Whereas some of them make calculations on the basis of a variance–covariance matrix, others calculate the reliability coefficient on the basis of factor analysis. Armor's θ and Heise and Bohrnstedt's Ω coefficients to be analyzed in this study are calculated according to exploratory factor analysis (EFA). McDonald ω, McDonald $ω_h$, and Maximal reliability can, on the other hand, be calculated according to confirmatory factor analysis (CFA). Revelle β and standardized alpha coefficients are also calculated on the basis of correlation matrices.

On reviewing the studies regarding reliability coefficients, no studies were found in the literature considering the reliability coefficients considered in this study altogether. Moreover, no studies were found analyzing the change of all these reliability coefficients with different EFA factor extraction methods and with CFA estimation methods. Therefore, it is believed that this study will contribute to the relevant literature.

This study is important in that it compares factor analysis methods and reliability values obtained through different reliability-determining techniques according to factor analysis results. Comparisons of the obtained reliability values that will demonstrate the similarities and differences in various conditions are also important for the experts of the field to plan and interpret prospective studies. In this way, individuals developing tests and using the results of tests will be able to decide appropriately on what technique is appropriate for use with which factor analysis method, with which sample size, and with which correlation technique. In line with the described purpose and importance of the study, answers are sought to the following questions:

- How do reliability coefficients calculated with CFA methods change according to sample size, correlation

matrices used, and factor extraction method?

- How do reliability coefficients calculated with EFA methods change according to sample size, correlation matrices used, and factor extraction method?
- How do reliability coefficients calculated with a correlation matrix change according to sample size and correlation matrices used?

2. Method

Since this study aiming to compare reliability coefficients is concerned with revealing the processes underlying a theory, it is a basic study. Researchers interested in basic studies are not particularly concerned with the results of educational applications. Instead, they develop the processes of a theory or analyze the processes (Fraenkel, Wallen, & Huyn, 2012; Gay, Mills, & Airasian, 2012).

Data

The data coming from a Turkish sub-test of the TEOG exam (test for transition from basic education to secondary education) given in the second semester of the 2013–2014 academic year constituted the data set of the research. A sample was taken from randomly chosen students answering the Turkish test to form the data set of 10,000 students. The response patterns that were repetitive or those showing that the students marked the same answer for all items were deleted from the data set, and thus a data set of 9,773 people was formed. The data set was then divided randomly into groups of 250, 500, 1,000, 2,000, and 5,000, ending up with the whole data set of 9,773. The data sets were analyzed in the form of item scores matrices scored between 1 and 0. A data set of 250 people, for instance, was formed once, and all coefficients were calculated by using this data set.

Analysis

In the current study, Armor θ, Heise and Bohrnstedt's Ω, McDonald ω, McDonald ω_h, maximal reliability, Revelle's β, and standardized alpha coefficient were investigated. Because Armor's θ coefficient requires the parallel measurement assumption of Cronbach's alpha coefficient, it is calculated by using principal components factor analysis as an alternative. Thus, Armor's θ coefficient is stated as

$$\theta = \frac{p}{p-1} \cdot \left(1 - \frac{1}{\lambda_1}\right) \qquad (1.3)$$

(Armor, 1974). Here, p represents the number of items, whereas λ_1 represents the biggest eigenvalue.

Heise and Bohrnstedt's Ω coefficient is obtained with EFA, as in the case of Armor's θ coefficient. Heise and Bohrnstedt's Ω coefficient is stated as

$$\Omega = 1 - \frac{\sum_{i=1}^{n} \sigma_i^2 - \sum_{i=1}^{n} \sigma_i^2 h_i^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} Cov(x_i, x_j)} \qquad (1.4)$$

(Heise & Bohrnstedt, 1970). Here, h_i^2 represents the communalities. Carmines and Zeller (1979) demonstrated that Heise and Bohrnstedt's Ω coefficient can also be represented as

$$\Omega = 1 - \frac{a - \sum_{i=1}^{n} h_i^2}{a + 2b} \qquad (1.5)$$

Here, *a* shows the number of items and *b* shows the sum of correlations between items.

McDonald's ω coefficient is obtained with CFA. If each unique variance is supposed to be composed of errors, then ω is equal to the reliability of the total score. McDonald's ω coefficient is stated as

$$\omega = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2}{\left(\sum_{i=1}^{k} \lambda_i\right)^2 + \sum_{i=1}^{k} \varphi_i} \qquad (1.6)$$

(McDonald, 1999). Here, λ_i represents the unstandardized factor loadings obtained with CFA, and φ_i represents unique variance.

McDonald's ω_h coefficient is calculated with λ_{gi} loads of the i. item on the general factor of McDonald's ω coefficient (Zinbarg, Revelle, Yovel, & Li, 2005). Accordingly, ω_h coefficient is stated as

$$\omega_h = \frac{\left(\sum_{i=1}^{k} \lambda_{gi}\right)^2}{Var(X)} \qquad (1.7)$$

Here, λ_{gi} represents the unstandardized factor load of i item to the factor.

Maximal reliability is stated as the reliability of a structure in congeneric measurements (Hancock & Mueller, 2001, 2013). Accordingly, maximal reliability is stated as

$$H = \frac{\sum_{i=1}^{k}\left(\frac{l_i^2}{1-l_i^2}\right)}{1+\sum_{i=1}^{k}\left(\frac{l_i^2}{1-l_i^2}\right)} \qquad (1.8)$$

Here, l_i^2 represents the square of standardized factor loadings.

Revelle's β coefficient is generally defined with the averages of item covariance of unequal length test halves (Revelle, 1979). This represents the average covariance of the items available in the different halves (Zinbarg et al., 2005). Revelle's β coefficient is stated as

$$\beta = k^2 \frac{\overline{\sigma_{ij}}}{Var(X)} \qquad (1.9)$$

Here, k represents the number of items, and $\overline{\sigma_{ij}}$ represents the covariance average between items.

The standardized alpha coefficient is standardized before the observed scores of the components are summed. Standardized alpha is not the lower limit of real reliability. It may be lower or higher than real reliability (Osburn, 2000). The standardized alpha coefficient is stated as

$$\text{Standardized } \alpha = \frac{n.\overline{p}}{1+(n-1).\overline{p}} \qquad (1.10)$$

(Carmines & Zeller, 1987; Osburn, 2000). Here, $\overline{p}$ stands for the correlation of the average between items, and n stands for the number of items.

Factor extraction methods were set as a condition for reliability coefficients calculated through EFA (Armor θ, Heise and Bohrnstedt's Ω) in this study. The methods of principal components analysis (PCA), principle axis (PA), unweighted least squares (ULS), weighted least squares (WLS), and maximum likelihood (MLE) were used for this purpose.

Estimation methods were set as a condition for reliability coefficients calculated through CFA (McDonald ω, McDonald ω_h, maximal reliability). For this purpose, reliability coefficients were calculated separately with the methods of ULS, ML, and generalized least squares (GL).

A Pearson correlation matrix was created as phi and tetrachoric for reliability coefficients, which were calculated with a covariance/correlation matrix, and calculations were made through them. A covariance matrix was also considered as a condition for reliability coefficients calculated through EFA and CFA, and reliability coefficients were calculated separately with Pearson, phi, and tetrachoric covariance matrices.

This study considers a sample size as six conditions (250, 500, 1,000, 2,500, 5,000, and 9,773), factor extraction for EFA as five conditions (PCA, PA, ULS, WLS, and MLE), estimation method for CFA as three conditions (UL, ML, and GL), and correlation/covariance matrix as three conditions (Pearson, phi, and tetrachoric). Accordingly, 6x3x5=90 conditions for each reliability coefficient calculated with EFA, 6x3x3=54 conditions for each reliability coefficient cal-

culated with CFA, and 6x3=18 conditions for each reliability coefficient calculated with a correlation/covariance matrix were set.

The samples were chosen randomly from the whole data set (9,773 students) through R program (R Core Team, 2017) in forming the data sets in this study. Tetrachoric correlations were calculated through the psych package (Revelle, 2016) in the R program in doing the analyses by means of correlation matrices with the data set formed. The Polycor package (Fox, 2016) was used for the Pearson correlation matrix. Function was described for a phi correlation matrix, and thus a phi correlation matrix was formed. Having formed the correlation matrices, the psych package was used for EFA and the Lisrel program was used for CFA. Calculations were made with the help of the formulas given in equations 1–7 by using the Excel program included in the Office 2016 package program by using the results obtained from the programs.

The needed restrictions were made with CFA so as to find which CTT measurement models the data set fitted. Evaluations were made as to which of parallel, tau equivalent, essentially tau equivalent, and congeneric measurement models fitted better, and it was found that congeneric models fitted better.

3. Findings

Table 1 shows all of the coefficients calculated in consequence. Interpretations concerning the sub-problems are also included.

Table 1. Comparison of Reliability Coefficients

Column groups — *Calculate via Confirmatory Factor Analysis*: McDonald ω, McDonald ωh, Maximal Reliability, PCA · *Calculate via Exploratory Factor Analysis*: Armor Θ, Heise ve Bohrnstedt'in Ω · *Calculate via Corr/Cov. Matrix*: Reveile β, Std. Alfa

Correlation Matrix	Sample Size	McD ω UL	McD ω ML	McD ω GL	McD ωh UL	McD ωh ML	McD ωh GL	Max Rel UL	Max Rel ML	Max Rel GL	PCA	Armor Θ PA	Armor Θ ULS	Armor Θ WLS	Armor Θ MLE	Heise Ω PCA	Heise Ω PA	Heise Ω ULS	Heise Ω WLS	Heise Ω MLE	Reveile β	Std. Alfa
Pearson	250	0.8789	0.8781	0.9018	0.8736	0.8701	0.8865	0.8883	0.8898	0.9146	0.8850	0.8639	0.8639	0.8639	0.8639	0.8890	0.8834	0.8834	0.8834	0.8834	0.8728	0.8824
Pearson	500	0.8662	0.8657	0.8788	0.8656	0.8643	0.8766	0.8730	0.8738	0.8862	0.8686	0.8417	0.8417	0.8417	0.8417	0.8728	0.8663	0.8663	0.8663	0.8663	0.8634	0.8652
Pearson	1000	0.8750	0.8744	0.8824	0.8742	0.8722	0.8789	0.8827	0.8835	0.8920	0.8793	0.8563	0.8563	0.8563	0.8563	0.8838	0.8778	0.8778	0.8778	0.8778	0.8735	0.8769
Pearson	2500	0.8776	0.8768	0.8830	0.8769	0.8735	0.8792	0.8840	0.8852	0.8912	0.8815	0.8591	0.8591	0.8591	0.8591	0.8862	0.8803	0.8803	0.8803	0.8803	0.8763	0.8796
Pearson	5000	0.8758	0.8751	0.8808	0.8749	0.8722	0.8775	0.8821	0.8831	0.8889	0.8800	0.8570	0.8570	0.8570	0.8570	0.8848	0.8788	0.8788	0.8788	0.8788	0.8745	0.8782
Pearson	All (9773)	0.8756	0.8749	0.8798	0.8750	0.8721	0.8766	0.8821	0.8832	0.8881	0.8799	0.8569	0.8569	0.8569	0.8569	0.8846	0.8787	0.8787	0.8787	0.8787	0.8743	0.8780
Phi	250	0.8790	0.8781	0.9017	0.8739	0.8703	0.8868	0.8885	0.8899	0.9144	0.8851	0.8640	0.8640	0.8640	0.8640	0.8891	0.8834	0.8834	0.8834	0.8834	0.8731	0.8824
Phi	500	0.8661	0.8657	0.8788	0.8653	0.8640	0.8765	0.8729	0.8737	0.8862	0.8685	0.8416	0.8416	0.8416	0.8416	0.8728	0.8663	0.8663	0.8663	0.8663	0.8631	0.8651
Phi	1000	0.8751	0.8745	0.8826	0.8745	0.8724	0.8794	0.8827	0.8836	0.8921	0.8794	0.8564	0.8564	0.8564	0.8564	0.8838	0.8779	0.8779	0.8779	0.8779	0.8740	0.8770
Phi	2500	0.8775	0.8768	0.8831	0.8766	0.8731	0.8790	0.8839	0.8851	0.8912	0.8815	0.8591	0.8591	0.8591	0.8591	0.8862	0.8803	0.8803	0.8803	0.8803	0.8759	0.8796
Phi	5000	0.8758	0.8751	0.8810	0.8821	0.8832	0.8881	0.8821	0.8832	0.8892	0.8800	0.8570	0.8570	0.8570	0.8570	0.8848	0.8789	0.8789	0.8789	0.8789	0.8746	0.8782
Phi	All (9773)	0.8754	0.8747	0.8796	0.8734	0.8707	0.8752	0.8819	0.8830	0.8879	0.8797	0.8566	0.8566	0.8566	0.8566	0.8844	0.8785	0.8785	0.8785	0.8785	0.8729	0.8777
Tetrakonik	250	0.9405	0.9396	N.C.	1.4456	1.4364	N.C	0.9508	0.9520	N.C	0.9452	0.9394	0.9397	0.9397	0.9394	0.9471	0.9445	0.9471	0.9471	0.9445	1.4436	0.9435
Tetrakonik	500	0.9343	0.9335	0.9568	1.4692	1.4587	1.4784	0.9418	0.9429	0.9655	0.9373	0.9301	0.9301	0.9301	0.9301	0.9394	0.9364	0.9364	0.9364	0.9364	1.4655	0.9354
Tetrakonik	1000	0.9392	0.9383	0.9517	1.4673	1.4572	1.4636	0.9477	0.9488	0.9632	0.9430	0.9368	0.9368	0.9368	0.9368	0.9451	0.9423	0.9423	0.9424	0.9423	1.4663	0.9414
Tetrakonik	2500	0.9406	0.9397	0.9493	1.4677	1.4569	1.4704	0.9482	0.9496	0.9576	0.9444	0.9384	0.9384	0.9384	0.9384	0.9466	0.9439	0.9438	0.9439	0.9438	1.4667	0.9431
Tetrakonik	5000	0.9398	0.9389	0.9469	1.4717	1.4606	1.4696	0.9473	0.9487	0.9566	0.9439	0.9377	0.9377	0.9377	0.9377	0.9460	0.9433	0.9433	0.9433	0.9433	1.4708	0.9426
Tetrakonik	All (9773)	0.9397	0.9389	0.9454	1.4717	1.4619	1.4705	0.9473	0.9487	0.9545	0.9438	0.9376	0.9376	0.9376	0.9376	0.9459	0.9432	0.9432	0.9433	0.9432	1.4709	0.9425

Note: N.C.: Not Converge

Analysis of Reliability Coefficients Calculated with CFA (McDonald ω, McDonald ω$_h$, and Maximal Reliability)

Table 1 shows reliability coefficients calculated according to such variables as sample size, estimation method, and correlation/covariance matrix. A review of reliability coefficients shows that reliability coefficients can be bigger than 1 when a tetrachoric covariance/correlation matrix is used. Thus, it was observed that the ω$_h$ and β coefficients took on values bigger than 1 when calculated with a tetrachoric covariance matrix. Besides, there is no convergence for the GL method in the sample of 250 students with a tetrachoric covariance matrix. On taking the reliability coefficient, which is calculated with the whole data set as a reference in analyzing the change of reliability coefficients according to sample size, the differentiation in the reliability coefficients obtained from the other samples starts after the third decimal place. Accordingly, it may be said that there is not much difference in the condition that the sample size is at least 250. When reliability coefficients are calculated with a Pearson and phi correlation/covariance matrix, the result is smaller than a tetrachoric correlation/covariance. A tetrachoric covariance matrix is used and recommended when the condition of normal distribution is satisfied for binary data (Allen & Yen, 1979; Price, 2017; Skrondal & Rabe-Hesketh, 2004). The average skewness in the data set was -1, and average kurtosis was about 0. Accordingly, the data can be said to have normal distribution (Chou & Bentler, 1995; Curran, West, & Finch, 1996). Even though the data had normal distribution, the ω$_h$ and β coefficients calculated with a tetrachoric correlation/covariance were observed to be bigger than 1.

When calculated from the data sets of 250, 2,500, and 5,000 students, the McDonald ω coefficient was higher than the one calculated from the whole data set, and the ω coefficient calculated from the data sets of 500 and 1,000 students was lower than the one calculated with the whole data set. However, the biggest difference was 0.0095, whereas the smallest was 0.0006. Accordingly, it may be said that the sample size in this study did not have serious effects on reliability coefficients. On comparing according to CFA estimation methods, the ω coefficient was found to be the lowest with the ML method and the highest with the GL method. Yet, the differences were small according to estimation methods, as in the case of sample sizes.

On evaluating with covariance matrices and CFA estimation methods, the McDonald ω coefficient took on the smallest value of 0.8657 and the biggest value of 0.9568. While the smallest value was observed in the sample of 500 students with a phi covariance matrix and with the ML method of estimation, the biggest value was observed in the sample of 500 students with a tetrachoric variance matrix and with the GL method of estimation.

The effects of sample size on coefficients were also slight for the McDonald ω$_h$ coefficient. On comparing according to CFA methods of estimation, it was found that the lowest estimation was made with the ML method and the highest with the GL method. The differences were also small according to estimation methods.

On evaluating the McDonald ω$_h$ coefficient along with all sample sizes, covariance matrices, and CFA estimation methods, it was found to take on the smallest value of 0.8640 and the biggest value of 1.4784. It took on the smallest value in the sample of 500 students with a phi covariance matrix and the ML method of estimation, whereas it took on the biggest value in the sample of 500 students with a tetrachoric covariance matrix and the GL method of estimation. Coefficient ω$_h$ took on values bigger than 1 in all samples and methods when calculated with a tetrachoric covariance matrix.

A comparison of maximal reliability coefficients according to CFA estimation methods showed the lowest values with the UL method and the highest values with the GL method. However, the differences were small. The biggest difference, which was approximately 0.02, was observed in the sample of 250 students.

On evaluating the maximal reliability coefficient with all sample sizes, covariance matrices, and CFA methods of estimation, it was found to take on the smallest value of 0.8729 and the biggest value of 0.9655. Whereas it took on the smallest value in the sample of 500 students with a phi covariance matrix and the UL method, it took on the biggest value in the sample of 500 students with a tetrachoric covariance matrix and the GL method.

Analysis of Reliability Coefficients Calculated with EFA (Armor's θ and Heise and Bohrnstedt's Ω)

The examination of Armor's θ coefficient and Heise and Bohrnstedt's Ω coefficients shown in Table 1 makes it clear that the reliability coefficient will be higher when PCA is chosen as the method of factor extraction. In addition, the other methods of factor extraction remained the same in the fourth decimal place, but they began to differ after the fourth decimal place.

Coefficient θ differs slightly according to factor extraction methods in the sample of 250 students when analyzed with a tetrachoric correlation matrix, but it is still the highest with the PCA method. When analyzed with Pearson and

phi correlation matrices, however, the PCA method yielded a θ coefficient higher than with the other methods. No big differences were found according to sample size.

On evaluating the θ coefficient along with all sample sizes, correlation matrices, and EFA factor extraction methods, it was found to take on the smallest value of 0.8416 and the biggest value of 0.9452. It took the smallest value in the sample of 500 students with a phi covariance matrix and the MLE method, whereas it took the biggest value in the sample of 250 students with a tetrachoric correlation matrix and the PCA method. The θ coefficient calculated with the MLE method is lower than the ones calculated with other methods. Yet, the difference is observed in the sixth decimal place.

When analyzed with a tetrachoric correlation matrix in a sample of 250 students, the Ω coefficient differs slightly according to factor extraction methods, but the highest values are observed with the PCA method. When analyzed with Pearson and phi correlation matrices, on the other hand, the PCA method yielded a Ω coefficient higher than with the other methods. No big differences were found according to sample size.

On evaluating the Ω coefficient along with all sample sizes, correlation matrices, and EFA factor extraction methods, it was found to take on the smallest value of 0.8663 and the biggest value of 0.9471. While it took on the smallest value in the sample of 500 students with a phi covariance matrix and the MLE method of factor extraction, it took on the biggest value in the sample of 250 students with a tetrachoric correlation matrix and the PCA method. The PCA, ULS, and WLS methods differ in the fifth decimal place. The Ω coefficient calculated with the MLE method was lower than with the other methods, but the difference was observed in the sixth decimal place.

On comparing the θ and Ω coefficients, it was found that the θ coefficient was lower than the Ω coefficient in all conditions. The PA factor extraction method yielded a lower θ coefficient than the other methods in analyses performed for Pearson, phi, and tetrachoric correlation matrices. It was found that the difference began in the fifth decimal place.

Analysis of Reliability Coefficients Calculated with a Correlation Matrix (Revelle β and Standardized Alpha)

The examination of Revelle β and standardized alpha coefficients shown in Table 1 makes it clear that the β coefficient calculated with a tetrachoric correlation matrix is bigger than 1. Standardized alpha, on the other hand, did not differ much according to sample size. The standardized alpha coefficient is bigger when calculated with a tetrachoric correlation matrix.

On evaluating the Revelle β coefficient along with all sample sizes, covariance matrices, and CFA methods of estimation, it was found to take on the smallest value of 0.8631 and the biggest value of 1.4709. Whereas it took on the smallest value in the sample of 500 students with a phi correlation matrix, it took on the biggest value when calculating with a tetrachoric correlation matrix from the whole data set. When the reliability coefficient was calculated with a tetrachoric correlation matrix, the result was bigger than 1.

On evaluating the standardized alpha coefficient along with all sample sizes, covariance matrices, and CFA methods of estimation, it was found to take on the smallest value of 0.8651 and the biggest value of 0.9435. While it took on the smallest value in the sample of 500 students with a phi correlation matrix, it took on the biggest value in the sample of 250 students with a tetrachoric correlation matrix.

4. Conclusion and Recommendations

On evaluating all of the reliability coefficients considered in this study, it was found that Armor's θ coefficient was lower than the other reliability coefficients in all conditions. The θ coefficient was found to be 0.8415 when estimating with the MLE method in all conditions, and it had the smallest value. The reliability coefficient found to be the highest (excluding those bigger than 1) was the maximal reliability coefficient. Maximal reliability calculated by making estimations with the GL method was found to have the value of 0.9655 at the maximum in all conditions.

While maximal reliability calculated with the CFA method was similar in terms of McDonald ω and McDonald $ω_h$ coefficients, the McDonald ω coefficient yielded values bigger than 1. Zinbarg et al. (2005) point out that the $ω_h$ coefficient and ω coefficient take on equal values for one-dimensional and unequal factor loadings. However, they take on values bigger than 1 when calculated with a tetrachoric correlation matrix. In a similar vein, Revelle's β coefficient also yielded values bigger than 1 when tetrachoric correlations were used. Zinbarg et al. (2005) also state that the β coefficient will be lower than the ω coefficient. Yet, Revelle's β coefficient has values bigger than 1 when calculated with a tetrachoric correlation matrix.

On assessing in terms of the same correlation matrix, it may be said that reliability coefficients do not differ much

according to sample size. This is a finding consistent with the ones obtained by Ercan, Yazici, Sigirli, Ediz, & Kan (2007). It was stated that the θ coefficient calculated with alpha and PCA did not differ according to sample size. Yet, the smallest sample size in this study was 250. Therefore, it was considered that it did not differ according to sample size, since it satisfied the condition of sample size (Comrey, 1988; Gorsuch, 1974; Guadagnoli & Velicer, 1988; Streiner, 1994) for factor analysis.

It may be said that sample size does not differ considerably in evaluating reliability coefficients for the same correlation matrix. Therefore, it is thought that sample size is an ignorable variable for reliability coefficients in this study on the condition that a sample of at least 250 students is available. Because McDonald ω_h and Revelle β coefficients were bigger than 1 for a tetrachoric correlation matrix and because there was overestimation, it may be said that calculating these coefficients with phi correlations in congeneric measurements, where variables are assumed to have real binary structure, would be more appropriate. It may also be said that Armor's θ and Heise and Bohrnstedt's Ω coefficients are higher than the other methods only in PCA, and therefore it may be recommended that the coefficients be calculated with the PCA factor extraction method, which tries to maximize the explained variance and thus raises the eigenvalue, since there are no differences in the other methods. The analyses were performed with one-dimensional binary data and with a real data set fitting a congeneric measurement model in this study. Research studies in which the factor structure and measurement model are modified through simulation studies could be recommended to researchers. The reliability coefficients considered in this study could be compared in such different conditions as smaller samples and small correlation averages. The research could also be repeated with ordinal and multi-category data types, and thus probable changes could be analyzed.

5. References

AERA, APA, NCME, American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement In Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.

Armor, D. J. (1974). Theta reliability and factor scaling. *Sociological Methodology*, *5*, 17–50.

Carmines, E. G., & Zeller, R. A. (1987). *Reliability and validity assessment* (9th ed.). Beverly Hills, California: SAGE.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: SAGE.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*(5), 754–761. https://doi.org/10.1037/0022-006X.56.5.754

Crocker, L., & Algina, J. (2008). *Introduction of classical and modern test theory*. Ohio: Cengage Learning.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining Cronbach alpha, theta, omega reliability coefficients according to sample size. *Journal of Modern Applied Statistical Methods*, *6*(1), 291–303. Retrieved from http://digitalcommons.wayne.edu/jmasm/vol6/iss1/27

Fox, J. (2016). polycor: Polychoric and polyserial correlations. Retrieved from https://cran.r-project.org/package=polycor

Fraenkel, J. R., Wallen, N. E., & Huyn, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). New Jersey: Pearson.

Gorsuch, R. L. (1974). *Factor Analysis* (1st ed.). Toronto: W. B. Saunders Company.

Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275.

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

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In *Structural Equation Modeling: Present and Future: A Festschrift in Honor of Karl Jöreskog* (pp. 195–216).

Hancock, G. R., & Mueller, R. O. (2013). *Structural equating modelling: A second course. B. G. Tabachnick & L. S. Fidell (Orgs.), Using multivariate statistics* (2nd ed.). Charlotte, NC: Information Age Publishing.

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

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: LEA Publisher.

Meyer, J. P. (2010). *Reliability*. New York: Oxford University Press.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*(3), 343–355. https://doi.org/10.1037/1082-989X.5.3.343

Price, L. R. (2017). *Psychometric methods: Theory and practice*. New York, NY: The Guilford Press.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*, 57–74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois. Retrieved from https://cran.r-project.org/package=psych

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: Multilevel, longitudinal, and structural equation models*. New York: Chapman & Hall.

Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *Canadian Journal of Psychiatry*, *39*(3), 135–140.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7