International Journal of Assessment Tools in Education
2023, Vol. 10, No. 4, 709–722

https://doi.org/10.21449/ijate.1271693

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate **Research Article**

Comparison of cronbach's alpha and McDonald's omega for ordinal data: Are they different?

Fatih Orcan[1,*]

[1]Kahramanmaraş Sütçü İmam University, Faculty of Education, Department of Educational Sciences, Türkiye

ARTICLE HISTORY

Received: Mar. 27, 2023
Revised: Oct. 03, 2023
Accepted: Oct. 16, 2023

Keywords:
Cronbach's Alpha,
McDonald's Omega,
Reliability,
MonteCarloSEM,
Simulation.

Abstract: Among all, Cronbach's Alpha and McDonald's Omega are commonly used for reliability estimations. The alpha uses inter-item correlations while omega is based on a factor analysis result. This study uses simulated ordinal data sets to test whether the alpha and omega produce different estimates. Their performances were compared according to the sample size, number of items, and deviance from tau equivalence. Based on the result, the alpha and omega had similar results, except for the small sample size, the smaller number of items, and the low factor loading values. When there were 5 or more items in the scale and factor analysis which the omega was calculated from showed fit to the data set, using omega over alpha could be preferred. Also, as the number of items exceeds 5, the alpha and omega differences disappear. Since calculating the alpha is easier compared to the omega (omega requires fitting a factor model first) using alpha over omega can also be suggested. However, when the number of items and the correlations among the items were small, omega performed worse than alpha. Therefore, alpha should be used for the reliability estimations.

1. INTRODUCTION

One of the most critical steps for a scientific study is data collection. This is also critical for reliability. Thus, the data collection process is better planned in order to get information as reliably as possible. Reliability is a property of the data collected, not the scale instrument itself (Streiner, 2003). Therefore, the data collected from the same instrument could be reliable for one example and not for another. One way of getting reliability is doing the test-retest procedure. According to the test-retest method, the same test is administered to the same group of examinees twice at a time interval. The correlation between the two-test administration is called test-retest reliability. Another method, which is commonly used (Streiner, 2003; Vaske, et al., 2017), is Cronbach's alpha (α). The alpha is also known as internal consistency reliability since it uses inter-item correlations to calculate reliability. The higher the correlations more reliable the data. The alpha value is also a property of the data. When the data changes, the alpha will also change. For reliable data collection, the change in the alpha is expected to be small compared to previous ones. However, as Henson et al. (2001) pointed the alpha values for the same instrument (a.k.a., Internal failure scale) ranged between .51 and .82 (as cited in Streiner, 2003).

*CONTACT: Fatih Orçan fatihorcan@ksu.edu.tr Kahramanmaraş Sütçü İmam University, Faculty of Education, Department of Educational Sciences, Kahramanmaraş, Türkiye

e-ISSN: 2148-7456 /© IJATE 2023

Cronbach's alpha comes with a few assumptions (Kalkbrenner, 2023). First of all, the items under the scale should be unidimensional (Edwards et al., 2021) which means all of the items are related only to one latent construct. Second, the items should be normally distributed and continuous (Edwards et al., 2021). Also, the error terms of the items should not be correlated. That is, there should not be any other common variance among the items except the latent factor. Last but not least, the items in the scale should equally contribute to the latent factor, the essentially tau equivalence assumptions. In case the assumption is not satisfied the alpha values underestimate the true reliability (Edwards et al., 2021; Kalkbrenner, 2023).

There are more than 30 reliability calculation methods (Edwards et al., 2021). Besides Cronbach's alpha, McDonald's omega (ω) is one of them. The alpha and omega values showed the most accurate estimate of reliability (Edwards et al., 2021). The omega reliability does not have such assumptions as the alpha. Therefore, when Cronbach's alpha does not hold its assumptions, it is recommended not to use alpha but to use omega instead (Goodboy & Martin, 2020). Specifically, when the tau equivalence assumption does not hold, use of ω is suggested (Viladrich et al., 2017). It is because, McDonald's omega is robust to the violations of the assumptions (Goodboy & Martin, 2020; Kalkbrenner, 2023). In fact, the alpha and omega give equally good results if the assumptions are met (Edwards et al., 2021; Goodboy & Martin, 2020; Viladrich et al., 2017).

The omega estimations are based on confirmatory factor analysis (CFA). A CFA model fits the data first and then the omega is calculated based on the factor loadings and the error variances as given in the formula:

$$\omega = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_i}$$

where the λ_i represents the factor loadings for item i, and θ_i represents the error variance of the item.

Even though omega does not have such assumptions as the alpha, since omega is calculated after a CFA, anything affecting model-data fit for the CFA model also affects the omega value. For example, the sample size is a critical issue for a factor analysis. As the sample size gets lower, the model-data fits become problematic for factor analyses or even the model may not converge to a solution (Gagne & Hancock, 2006). Also, under small sample sizes and unequal factor loadings omega estimation becomes biased (Edwards et al., 2021). Also, the number of items in the structure affects the omega reliability. Increasing the number of items stabilizes the omega estimates even for small sample sizes (Edwards et al., 2021; Ercan et al., 2017). In short, even though ω is a good alternative to α, the alpha produced more accurate estimates under a small sample size and number of items (Edwards et al., 2021).

The alpha and omega estimates also differ when the factor loadings have different values (e.g., non-tau equivalence) under a factor analysis (Edwards et al., 2021). When there is a discrepancy among the factor loading and as the size of the discrepancy increases, the alpha and omega produce different results. However, the difference between alpha and omega has "no practical consequences" when the average factor loadings are .7 or higher and the difference among the loading values is .2 in absolute values (Raykow & Marcoulides; 2015; Viladrich et al., 2017).

1.1. Aim of the Study

Edwards et al. (2021) compared six different reliability estimations in their work and based on the results, alpha and omega produced the most accurate estimate of the true reliability. Even though the alpha and omega were shown to be better, it was also shown that each estimation has its flaws. For example, "omega was affected greater by the number of items when reliability was low" (p. 1111). However, the work of Edwards et al. (2021) was based on continuous scaled data. In applications, ordinal scaled data are often used. Seeing the performance of alpha and omega with ordinal scaled data is informative and, therefore, this paper aims to estimate alpha and omega reliabilities based on a five-point Likert-type ordinal scale as similar conditions with Edwards et al. (2021).

To compare the alpha and omega estimates data were simulated under different conditions: sample size, balanced and un-balanced factor loadings, and number of items.

2. METHOD

2.1. Data Generation Procedure

Data were simulated via *MonteCarloSEM* package (Orçan, 2021) in R-Cran (R Core Team, 2014). The *MonteCarloSEM* package simulates and analyzes data based on a given CFA model. It can produce normal/skewed or continuous/ordinal scale data sets for given threshold values. Various simulation conditions were considered in this study. First of all, the sample size varied from 50 to 1000 to represent the sample size from small to large (e.g., 50, 100, 300, 500, 800, 1000). Second, the number of items under the scale differed. The minimum number under the scale was 3 and increased to 5, 8, 10, and 20. Therefore, five different number of items were considered for this study. The minimum number of items was chosen to be 3 because it is a prerequisite for a single-factor CFA model. Also, the maximum was 20 since the correlation between the number of items and reliability was shown to be reduced after 19 items (Vaske et al., 2017). Finally, average factor loading values were also changed for this study. Under this condition, five different scenarios were tested.

- Tau: All factor loadings were equal across the factor and the average loadings were set to be .3, .4, .5, .6, .7, .8, and .9.
- Mixed-1: The loadings values were differed by .2 at maximum and average loadings were set to be .3, .4, .5, .6, .7, .8, and .9. Therefore, for the average of .3, the factor loadings were .2, .3 and .4 under three items and .2, .25, .3, .35, and .4 repeated each for four times under 20 items.
- Mixed-2: The loadings values were differed by .4 at maximum and the average loadings were set to be .4, .5, .6, and .7. For the average of .4, under three items, the loadings were .2, .4 and .6 and under twenty items the loadings were .2, .3, .4, .5, and .6 each repeated for four times.
- Mixed-3: The loadings values were differed by .5 at maximum and the average loadings were set to be .5, .6, and .7. For the average of .5, under three items, the loadings were .25, .5, and .75 and under twenty items the loadings were .25, .35, .5, .65, and .75 each repeated for four times.
- Mixed-4: The loadings values were differed by .6 at maximum and the average loadings were set to be .5 and .6. For the average of .5, under three items, the loadings were .2, .5, and .8 and under twenty items the loadings were .2, .3, .5, .7, and .8 each repeated for four times.

For all the scenarios, the loadings were increased by .1 each time to increase the average factor loadings, respectively. For example, under the mixed-4 condition, to get average loadings to be .6 for three items the loadings were set as .3, .6, and .9.

2.2. Data Analysis

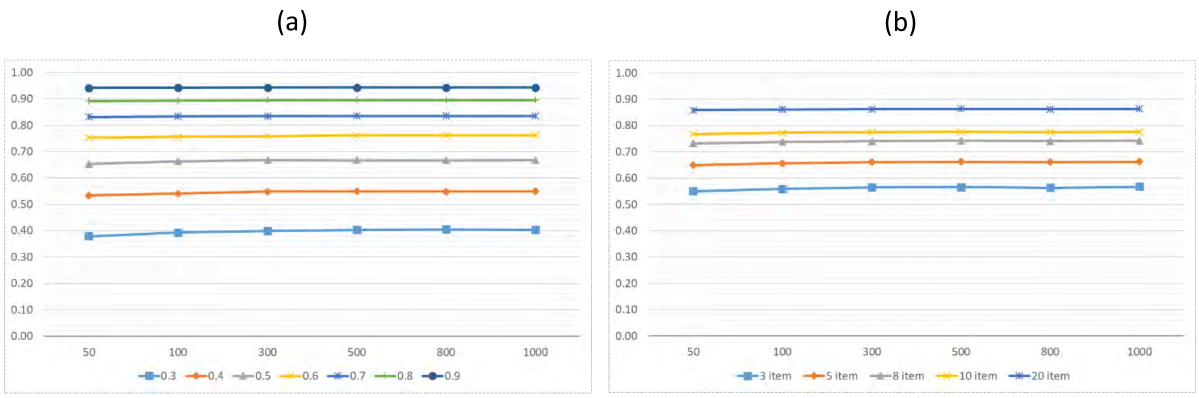
Each condition was repeated 1000 times by using the *sim.categoric()* function in the *MonteCarloSEM* package. Ordinal data were created by using -1.645, -.643, .643, and 1.645 as the threshold values to create a 5-point Likert scale. The Cronbach's alpha values were calculated by using *CronbachAlpha()* function in the *DescTools* package (Signorell, 2023). To estimate omega values one-factor CFA models were fitted to the simulated data sets by the *cfa()* function in the *lavaan* package (Rosseel, 2012), using the maximum likelihood estimation method. Besides the omega, model-data convergence rates and fit indices such as the p-value of the chi-square test, the comparative fit indices (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) were saved for the further evaluations. Hu and Bentler's (1999) criteria were used to evaluate the fit. Moreover, the relative biases where the absolute difference between the true and estimated values were divided by the true value were calculated:

$$Relative\ Bias = \frac{Abs(True\ Value - Estimated\ Value)}{True\ Value}$$

3. RESULTS

Based on the results of the simulation, the sample size was not an important factor for the Cronbach alpha estimation, as expected. The average estimates were almost identical across different sample sizes. As the sample size was increased from 50 to 1000, the alpha estimates changed only by .02 for the low factor loading values (.3, .4, and .5). When the factor loading values were increased, the gap disappeared and the estimates became identical. Similarly, when the number of items was 3, the average alpha estimates differed by .02, and as the number of items got larger, the gab became .01. Figure 1 showed the change of average alpha estimates by sample size for different values of the factor loadings and the number of items. As expected, the estimates increased by the increase in the factor loadings and the number of items. However, the estimated values had horizontal lines across different sample sizes.

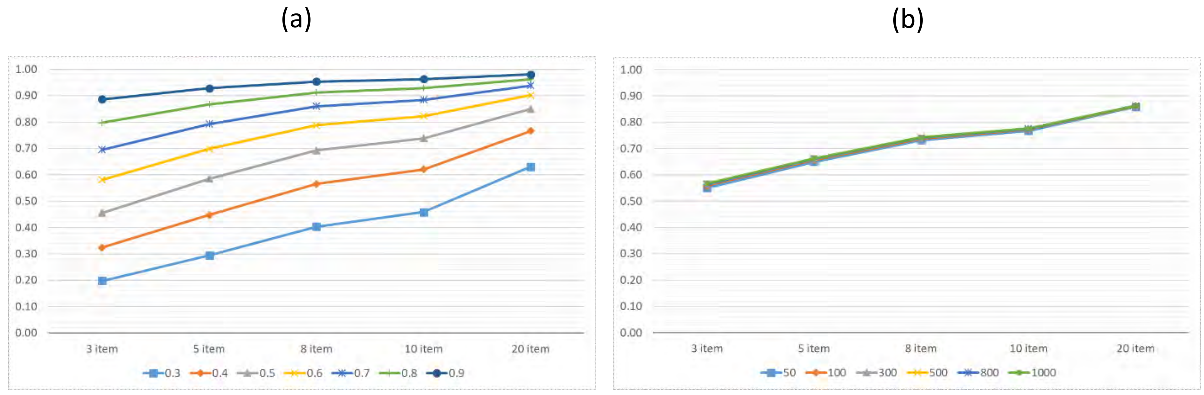
Figure 1. *Alpha estimated by sample size for factor loadings (a) and number of items (b).*



Different from the sample size, the number of items affected the alpha estimates according to the simulation results. Figure 2 shows the change of average alpha estimates by the number of items for different values of factor loadings and sample sizes. As the total number of items was increased, the estimates also increased. This was the result expected, since the effect of the number of items on the alpha is well-known in the literature (Streiner, 2003; Vaske et al., 2017). In addition, the effects of the number of items on the alpha estimates were larger, when the factor loadings were smaller (See Figure 2/a). For example, when the factor loading was .3 and the number of items was increased from 3 to 20, the estimate jumped to .63 from .20. However, when the factor loading was .6, the change in the estimates was smaller, from .58 to .90. The gap even got smaller when the factor loading was .9, from .89 to .98.

Based on the results of the simulation, under the model where the factor loadings were equal (Tau model), the α and ω values differed only for small sample sizes (50 and 100), small factor loadings (.3 and .4), and less number of items (3 and 5). The difference between the values ranged between .06 and .18. For all other conditions under the tau models difference was smaller than .04 and as the sample size, number of items, and factor loadings were increased, the gaps disappeared. Figure 3 shows the alpha (α) and omega (ω) estimates as well as the relative biases and the true values under the tau models for sample sizes of 50, 300, and 1000 (only three sample sizes were given due to space limitations).

Figure 2. *Alpha estimated by number of items for factor loadings (a) and sample sizes (b).*



For example, when there were 3 items, the loadings were all set to be .3, and the sample size of 50, the alpha and omega estimates were .18 and .35, respectively where the true value was .23. However, when holding all other values constant but the sample size was increased to 100, the values become .19 and .33. On the other hand, when the number of items becomes 5 instead of 3, the values become .27 and .39, indicating a .12 difference. Similarly, when factor loadings were increased to .4, the values became .31 and .45. Besides, the omega estimates were always larger than the alpha, except for sample sizes of 50, 20 items, and factor loadings of .3. Under this condition, the alpha and omega values were .62 and .61, respectively.

Figure 3 also showed relative biases for α and w. Based on Bandalos's (2002) recommendation, which pointed that relative bias should be smaller than .10, the relative biases pointed to problematic values under small sample size (50 and 100), small factor loadings (.3 and .4), or a smaller number of items (3 and 5). That is, as the sample size, number of items, and/or average factor loadings were increased, the relative biases decreased and got under the .10 critical value. Interestingly, under these conditions, alpha and omega estimates pointed to almost identical relative biases. For example, the relative biases showed similar values even when the sample size was as small as 50, the number of items 8, and factor loadings .4. Increasing values of any of these simulation conditions diminished the relative biases and the gap between the alpha and omega estimates.

Almost identical results with the Tau model were obtained for the mixed factor loadings model where the factor loading differs only by .2 (Mixed-1 model). Therefore, when tau equivalence was not granted and the difference between the loadings was up to .2, using omega reliability instead of the alpha did not change the results, except for small sample size (50 and 100), small factor loadings (.3 and .4) and less number of items (3 and 5) as it was the case for the tau models. That is to say, even for a small sample size the alpha and omega were almost identical (the difference was at the third decimal) as long as the factor loading differs by .2, the average factor loadings were above .5, and the number of items was more than 5. Figure 4 shows the alpha (α) and omega (ω) estimates for Mixed-1 models for the sample sizes of 50, 300, and 1000. Based on Mixed-1 model results, when the sample size was increased, the gap between the alpha and omega estimated disappeared even for small factor loadings and/or fewer items. For example, when the sample size was 300, the number of items was 3, and the average factor loading was .5, the difference between alpha and omega was only .02. Keeping everything constant but increasing the number of items to 5, the difference disappeared (.006).

Figure 3. *Alpha and omega estimates and relative biases for tau models under different sample sizes.*

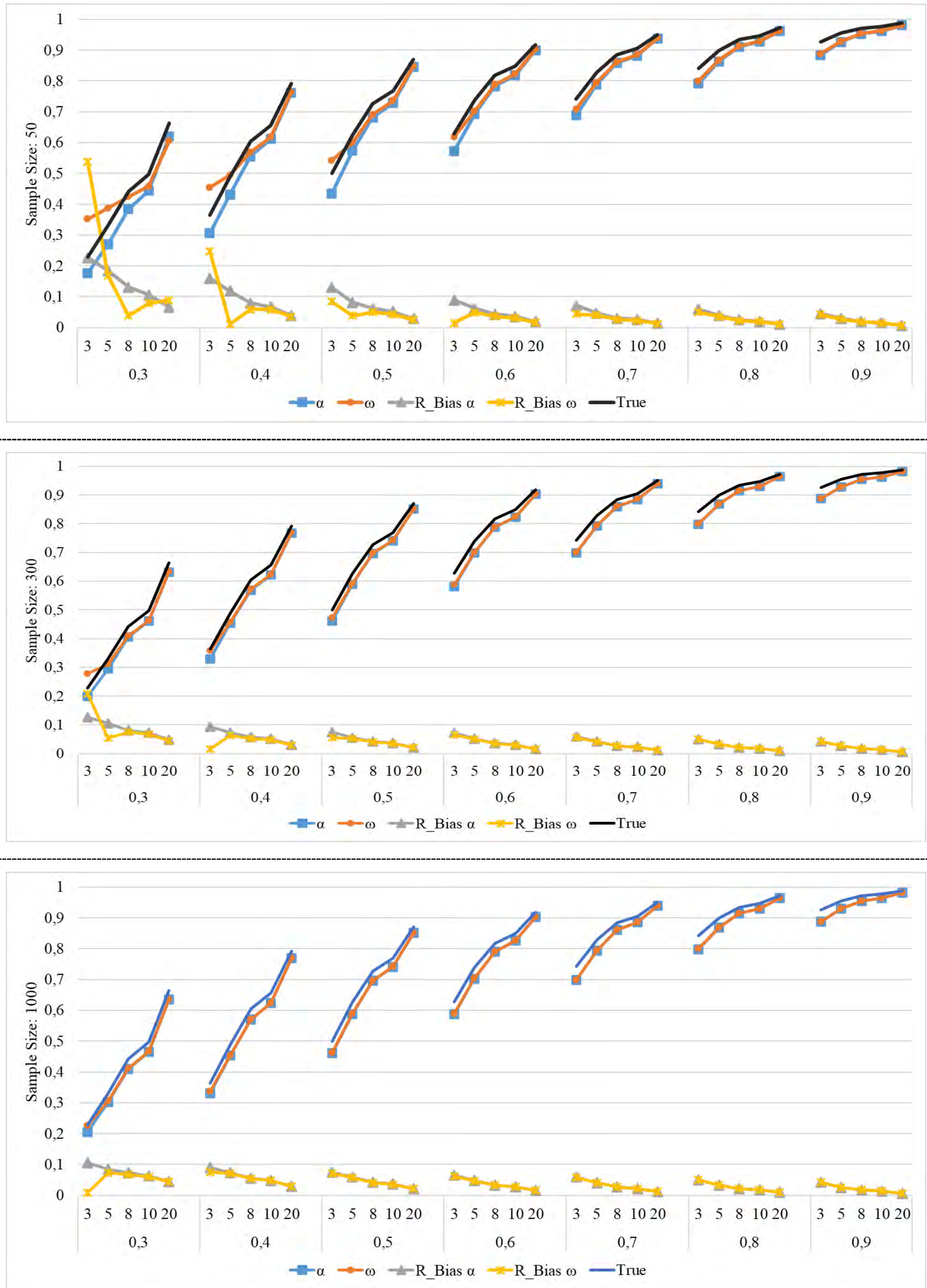


Figure 4. *Alpha and omega estimates and relative biases for mixed-1 models under different sample sizes.*

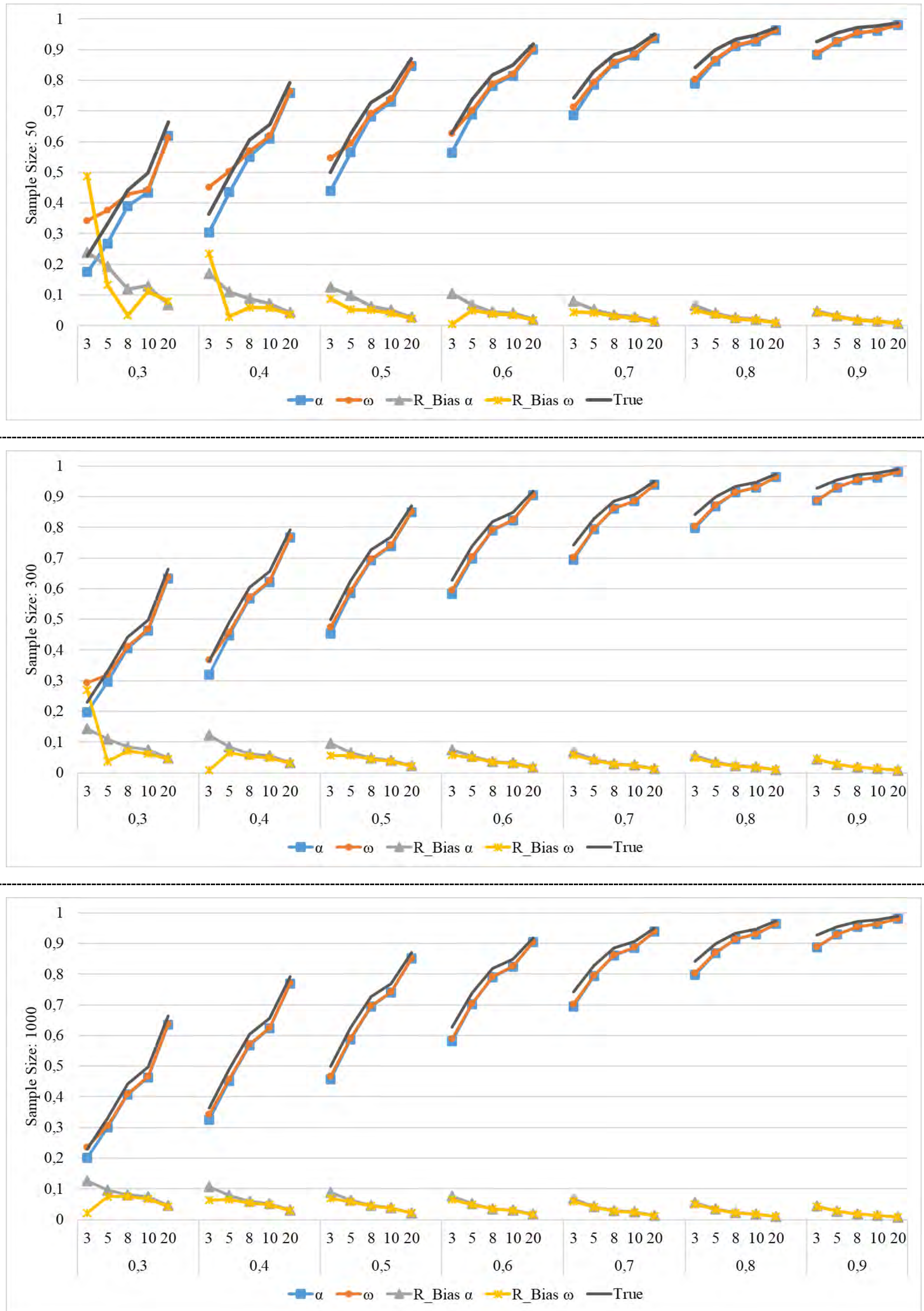


Figure 5. *Alpha and omega estimates and relative biases for mixed-2 models under different sample sizes.*

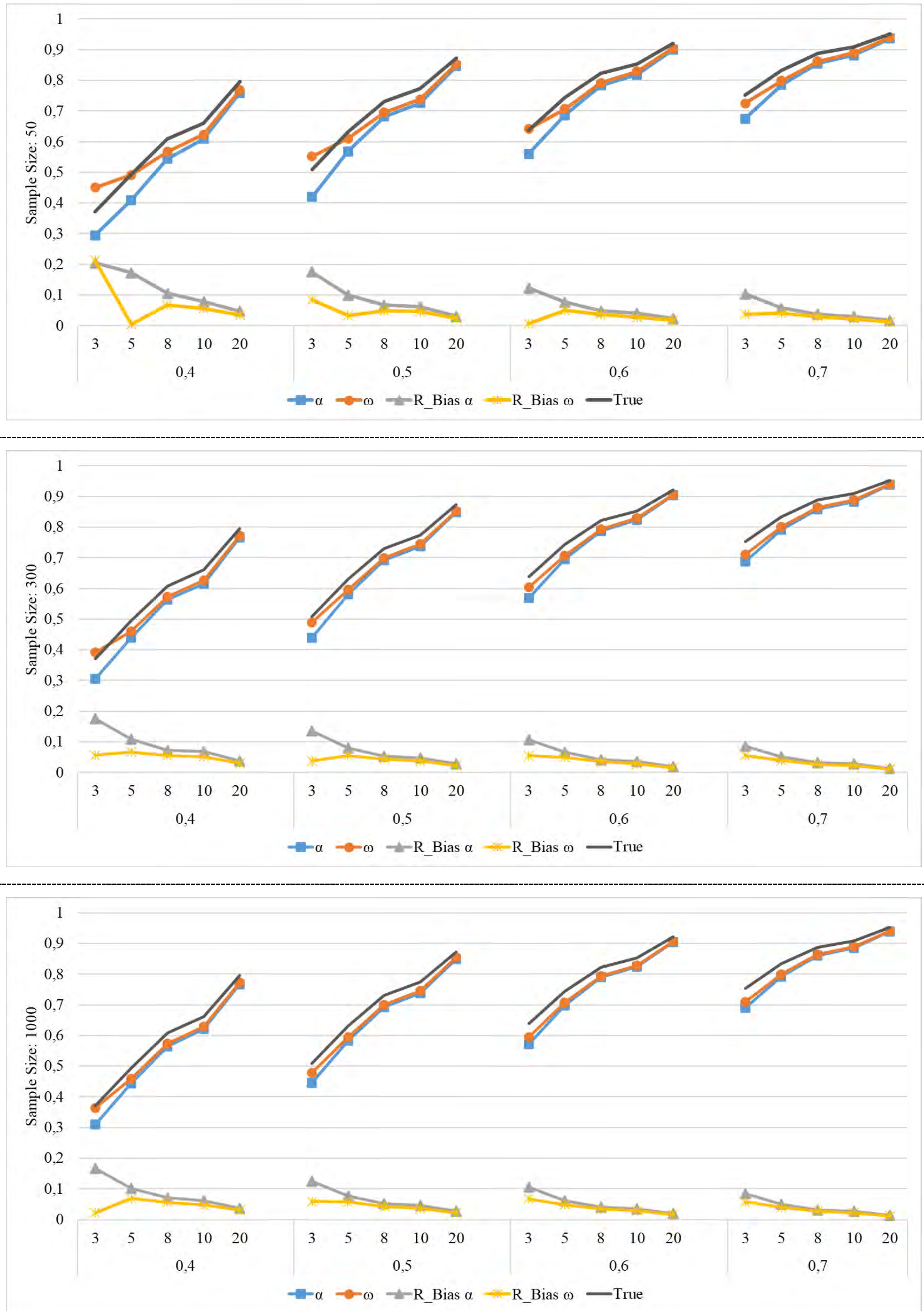
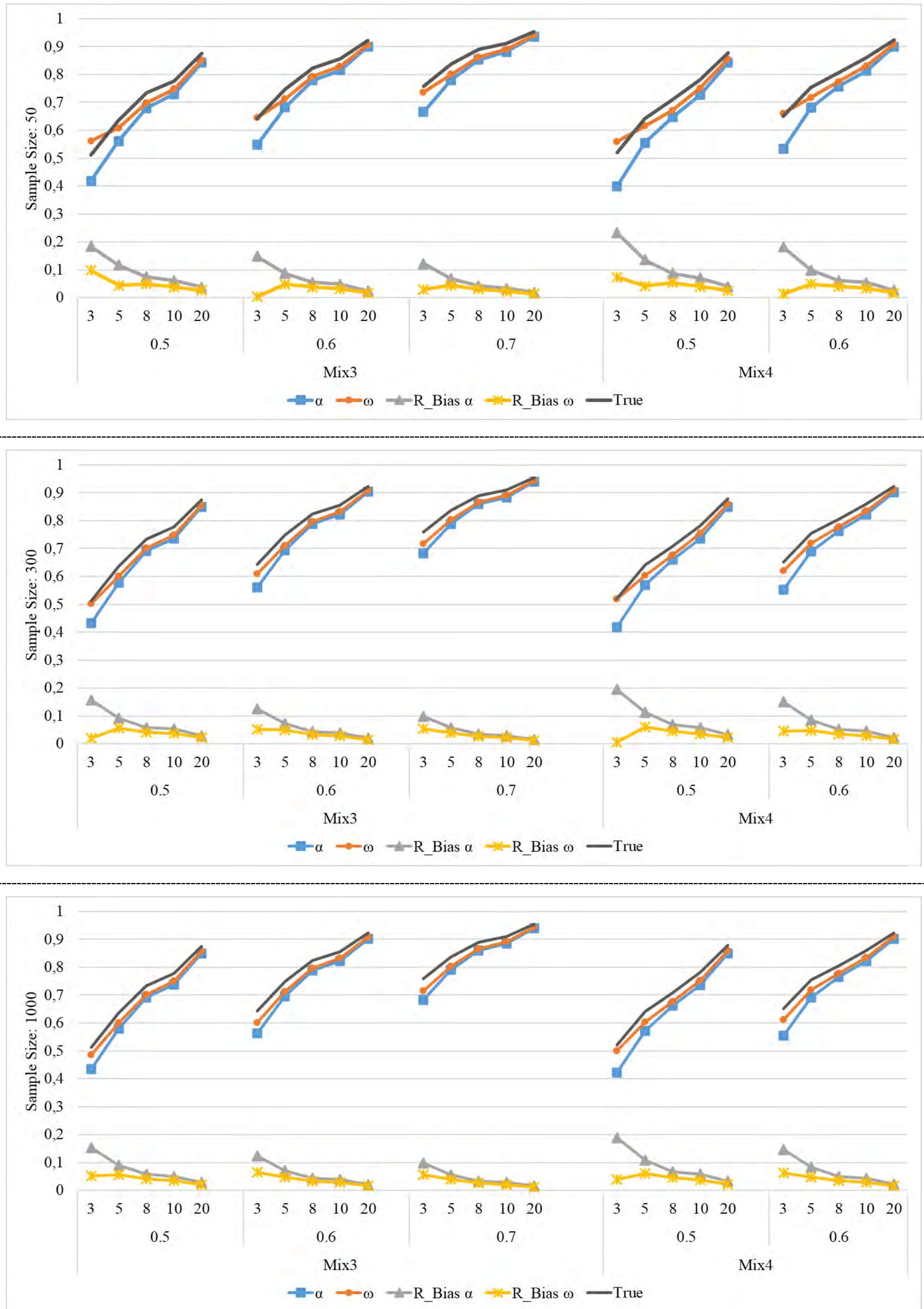


Figure 6. *Alpha and omega estimates and relative biases for mixed-3 and mixed-4 models under different sample sizes.*



Similar results were produced among the other mixed factor loadings (a.k.a., non-tau) models where the factor loading differs by .4 (Mixed-2 model), the factor loading differs by .5 (Mixed-3 model) and the factor loading differs by .6 (Mixed-4 model). The results of the models are reported in Figure 5 and Figure 6. When the number of items was 3 and average factor loadings were smaller than .6, all these conditions produced considerable gaps between the alpha and omega estimates, even for a sample size of 1000. That is, when there were 3 items at a scale and the average of factor loadings was smaller than .6, the tau equivalence became important under the Mixed-2, 3, and 4 models. Even though the gap got smaller with the sample size for some conditions, the change was smaller than .02.

Therefore, based on the results of the study, using alpha or omega to estimate reliability becomes not an important issue when the sample size and the number of items are larger than 300 and 3, respectively. The average of factor loadings and the difference between the factor loadings at a scale become important only when the number of items is 3.

The omega estimates were based on factor analysis results. Therefore, a CFA model should be tested, and the results should show a good model-data fit first. According to the simulation results, when the sample size and factor loadings were small, under 3 and 5 items models, the CFA models showed a higher percent of convergence problems. In other words, when there were more than 5 items, the average factor loadings were larger than .5, and the sample size was larger than 100, the model convergence was not a problem. Figure 7 shows the percentage of non-convergences under different sample sizes (Only the results of the sample sizes of 50 and 100 were given since as the sample size increased, they had no convergences). For example, when the sample size was 50, the number of items was 3, and the average factor loading was .3 under Tau and Mixed-1 models, about 28% of the data did not converge to a solution. The non-convergence rates decreased as the sample size, the number of items, and the average factor loadings were increased. Also, when the sample size and factor loadings were small, under 3 and 5 items models, the supplementary fit indices (CFI, RMSEA, and SRMR) indicated model-data fit issues too, in case the model converged to a result.

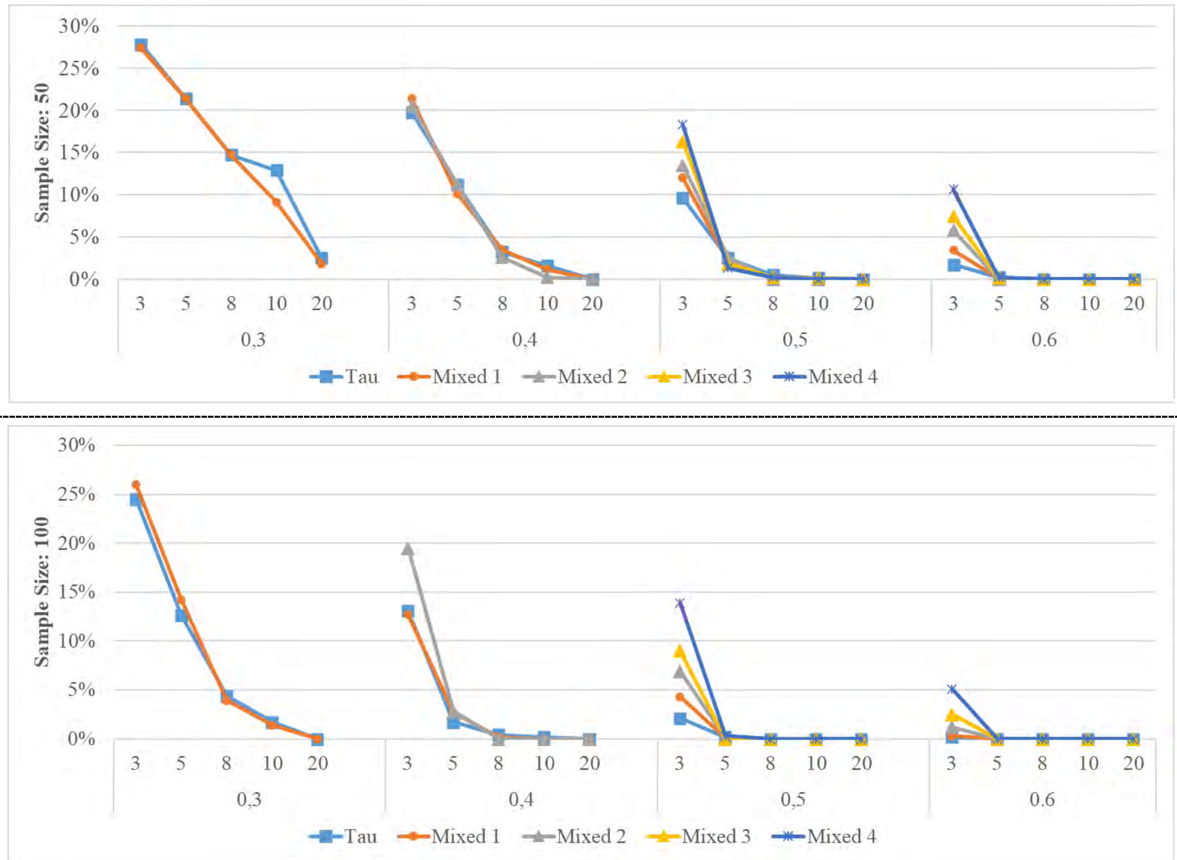
4. DISCUSSION and CONCLUSION

Among others, Cronbach's Alpha (α) and McDonalds' Omega (ω) were used commonly for the reliability estimates. Also, it was shown that the alpha and omega produced the most accurate and similar estimates of reliability (Edwards et al., 2021). Therefore, only these two reliability estimates were considered in this study. Based on the results of the current study, α and ω had similar results, except for the small sample size, a smaller number of items, and low factor loading values. Since the omega estimates were based on CFA results and factor analysis requires a larger sample size to converge a solution, under small sample sizes, the ratio of convergence was low. Therefore, the gap between the estimates of α and ω might be due to the convergence problem. Also, as the convergence rate increased with sample size, number of items, and factor loading, the gap between the estimates of α and ω got smaller and disappeared eventually.

Related to the convergence problem, even when the model converged to a solution, supplementary fit indices (CFI, RMSEA, and SRMR) sometimes indicated problems regarding the model-data fit. Similar to the convergence problem, the fit indices showed problems only when the sample size was small, the number of items and the factor loading values were low. Therefore, in case the data fit creates a concern, the estimated omega values might be problematic and "should not be used" to estimate reliability (McDonald, 2011, p. 89).

Relative biases were also calculated for the estimates (See Figures 3 to 6). When the number of items was larger than 5, relative biases were almost identical under all models (Tau, Mixed-1, 2, 3, and 4), regardless of the average factor loadings or the sample sizes.

Figure 7. *Percent of non-convergences under different sample sizes.*



When the number of items was 3, the sample size was small and the average factor loadings were .4 and smaller, omega showed greater bias than alpha. However, when the number of items was increased to 5, the bias of omega was smaller than that of alpha. Therefore, similar to the convergence issue, it can be concluded that when sample size, number of items, and average factor loadings were small, the omega estimates became worse than the alpha estimates. However, under all other conditions, ω produced less or equal relative biases than α. Therefore, when there are 5 or more items in the scale and the CFA model fits to the data set, using ω over α could be preferred since it produced less or equal bias compared to the α. It was also worth pointing that as the number of items becomes larger than 5 the difference between the estimates becomes smaller. Therefore, under these conditions, using alpha or omega to estimate reliability does not affect the results. From this point of view, since calculating the alpha is easier compared to omega because omega requires fitting a CFA model first, using alpha over omega can also be suggested.

Under mixed-2, mixed-3, and mixed-4 models, since average factor loadings were all larger than .4 due to the design factors, the omega only outperformed the alpha estimates when the number of items was 5 or smaller. Especially when the sample size and number of items were small, the alpha produced relative biases larger than the critical value (10%). If there were more than 5 items, the alpha and omega estimates were almost indistinguishable. To conclude, when there were more than 5 items, alpha and omega produced similar results, regardless of sample size, average factor loadings, or the tau equivalence assumptions. Therefore, based on the results, using alpha to estimate reliability is not wrong. However, when the CFA model, which the omega is calculated from, fits the data well, using omega to estimate reliability is also reasonable.

In conclusion:

- Even with tau models, alpha produced biased results under the small number of item number, sample size, and average factor loadings. However, under the same conditions, omega produced more bias. Similarly, Edwars et al. (2021) reported that omega was affected more by the sample size and average factor loadings. Even though it was not reported in the study, all the biases produced by alpha were positive, indicating alpha underestimates the true value while omega produced negative biases for the small sample sizes, fewer items, and low factor loadings.
- Based on the results of the study, it is true that the further away from the tau, the more biased the alpha is compared to the omega. Similar results were also reported by Edwards et al. (2021). However, as values of the design factors increase, regardless of the sample size, alpha and omega yield similar results. Thus, as long as the values of the design factors are not small, there is no harm in using the alpha for reliability.
- The results again confirmed that alpha under-estimate the reliability. However, it only happens as the model deviates from the tau equivalence and has a smaller number of items and average factor loadings. Under other conditions, the differences between alpha and omega estimates are less than 3%.
- To estimate omega, a CFA model is required to run first. If the CFA model does not fit to the data, the omega obtained from it may also be biased. However, this is not the case for the alpha estimates. Alpha gives a result even when the CFA produces problematic results for the omega estimates (e.g., smaller sample sizes, average factor loadings, and number of items). In fact, the bias of the alpha is even lower compared to omega under these conditions. Similarly, Edwards et al. (2021) have shown that alpha was superior to omega under these conditions.
- In the case of small sample size, number of items, and low factor loadings, omega estimates showed a much larger bias compared to the bias of alpha. As the values of the design factors were increased the bias of omega and alpha became closer. The larger omega bias was most probably due to the convergence problems of the CFA model. Since the models do not converge, the results were not stable and showed deviated estimates.
- When the number of items is 3 and the correlations among the items are low, even if the sample size is 1000 and the tau equivalence holds, the alpha estimates show biased results (over 10%). However, it should be remembered that there might be convergence and fit problems for the omega under similar conditions. Therefore, if there are low correlations among the items, increasing the number of items is important to get a more accurate reliability prediction.
- Even though, it seems that both α and ω produce biased results when the sample size is small, keeping the sample size constant and increasing the average factor loadings and the number of items, the bias disappears. Despite that Edwards et al. (2021) reported the effect of sample size on the omega estimates, this study showed that it is not directly related to the omega itself. It is most probably related to the requirements of CFA models. Since a CFA model needs a larger sample size for stable estimations, in case the sample size was low, the omega might not be estimated correctly. Therefore, it cannot be said that the sample size had a direct effect on the biases. After all, Yurdagül (2008) showed that α produces unbiased estimates even when the sample size is as low as 30. In a similar situation, the high bias of ω can be associated with the model data fits.

The results were limited to the simulation conditions. Under this study, one-factor CFA models were considered. Similar comparisons of reliability estimations can also be made for multi-dimensional (a.k.a., two or more factor CFA models) structures.

Declaration of Conflicting Interests and Ethics

The author declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

Orcid

Fatih Orçan https://orcid.org/0000-0003-1727-0456

REFERENCES

Bandalos, D.L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5

Bernardi, R.A. (1994). Validating research results when cronbach's alpha is below .70: A methodological procedure. *Educational and Psychological Measurement, 54*(3), 766–775. https://doi.org/10.1177/0013164494054003023

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Edwards, A.A., Joyner, K.J., & Schatschneider, C. (2021). A simulation study on the performance of different reliability estimation methods. *Educational and Psychological Measurement, 81*(6), 1089-1117. https://doi.org/10.1177/0013164421994184

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. https://doi.org/10.22237/jmasm/1177993560

Gagne, P., & Hancock, G.R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research, 41*(1), 65-83. https://doi.org/10.1207/s15327906mbr4101_5

Goodboy, A.K., & Martin, M.M. (2020). Omega over alpha for reliability estimation of unidimensional communication measures. *Annals of the International Communication Association, 44*(4), 422-439. https://doi.org/10.1080/23808985.2020.1846135

Henson, R.K., Kogan, L.R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement, 61*(3), 404-420. https://doi.org/10.1177/00131640121971284

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. https://doi.org/10.1080/10705519909540118

Kalkbrenner, M.T. (2023). Alpha, omega, and h internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation, 14*(1), 77-88. https://doi.org/10.1080/21501378.2021.1940118

McDonald, R.P. (2011). *Test theory: A unified treatment*. Routlege.

Orçan, F. (2021). MonteCarloSEM: An R package to simulate data for SEM. *International Journal of Assessment Tools in Education, 8*(3), 704-713. https://dergipark.org.tr/en/pub/ijate/issue/62753/804203

R Core Team. (2014). *R: A language and environment for statistical computing* [Computer software manual]. http://www.R-project.org/

Raykov, T., & Marcoulides, G.A. (2015). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and Psychological Measurement, 75*(1), 146–156. https://doi.org/10.1177/0013164414526039

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Signorell, A. (2023). DescTools: Tools for descriptive statistics. R package version 0.99.48. https://CRAN.R-project.org/package=DescTools

Streiner, D.L. (2003). Starting at the Beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103. https://doi.org/10.1207/S15327752JPA8001_18

Vaske, J.J., Beaman, J., & Sponarski, C.C. (2017). Rethinking internal consistency in cronbach's alpha. *Leisure Sciences, 39*(2), 163-173. https://doi.org/10.1080/01490400.2015.1127189

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología / Annals of Psychology, 33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401

Yurdagül, H., (2008). Minimum sample size for cronbach's coefficient alpha: A Monte-Carlo Study. *H.U. Journal of Education, 35*, 397-405. http://www.efdergi.hacettepe.edu.tr/shw_artcl-571.html