

Stability of the Data-Model Fit over Increasing Levels of Factorial Invariance for Different Features of Design in Factor Analysis

Deyab Almaleki

Department of Evaluation, Measurement, and Research
Umm Al-Qura University
Makkah, Saudi Arabia
damaleki@uqu.edu.sa

Abstract—The aim of this study is to provide an empirical evaluation of the influence of different aspects of design in the context of factor analysis in terms of model stability. The overall model stability of factor solutions was evaluated by the examination of the order for testing three levels of Measurement Invariance (MIV) starting with configural invariance (model 0). Model testing was evaluated by the Chi-square difference test ($\Delta\chi^2$) between two groups, and Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI). Factorial invariance results revealed that the stability of the models was varying over increasing levels of measurement as a function of Variable-To-Factor (VTF) ratio, Subject-To-Variable (STV) ratio, and their interactions. There were invariant factor loadings and invariant intercepts among the groups indicating that measurement invariance was achieved. For VTF ratios 4:1, 7:1, and 10:1, the models started to show stability over the levels of measurement when the STV ratio was 4:1. Yet, the frequency of stability models over 1000 replications increased (from 77% to 91%) as the STV ratio increased. The models showed more stability at or above 32:1 STV.

Keywords—model stability; factorial invariance; level of measurement invariance; model design

I. INTRODUCTION

Confirmatory Factor Analysis (CFA) is a form of Factor Analysis (FA) that tests hypotheses regarding how well the measured indicator variables represent the number of constructs [1]. CFA is a confirmatory method a researcher can use to examine, evaluate, and/or test a number of hypothesized factors underlying the variance/covariances in a set of measured indicator variables. CFA allows the researcher to test hypothetical and plausible alternative latent variable structures for the observed indicator variance/covariances [2]. More recently, CFA has also been used in exploratory analysis too [3-9]. Three major concerns have emerged repeatedly in the literature related to the use and interpretation of FA in social science research: (a) determining an adequate number of indicator variables to describe the latent trait, (b) factoring a sufficient sample size to have reasonable confidence in the stability of the model estimate, and (c) establishing minimum communality levels to determine which indicator variables can

represent a latent trait, especially in simulation studies [8, 10-13]. FA assumes that the indicator variables used should be linearly related to one another. Otherwise, the number of extracted factors will be the same as the number of original variables [2, 15]. Survey instrument length and the number of variables differ based on discipline, purpose, sample frame, and method of data collection. Recently, the online survey has become an important method of data collection for a variety of reasons (e.g. online surveys are easy to design, conduct, and often they are the only option for data collection). According to SurveyMonkey the median length of its paid surveys was 10 questions [9]. Industry-specific surveys and market-research surveys tend to have more questions while event surveys and just-for-fun surveys tend to be shorter (see Figure 1) [9]. If the length of the survey is about 10 questions or fewer, it can lead to a higher completion rate and increase the likelihood that people will choose to take more of the researcher's surveys in the future. More recent studies of factor analysis do not include the VTF ratio 10:1 in their investigations [5, 12, 14-17], nor the way this number is relative to the sample size or the communality magnitude when factor analysis is conducted.

II. LITERATURE REVIEW

A. Observed Variables

In FA, the observed indicator variables can be viewed as representing a sample of potential variables, all of which measure the same construct or factor [1, 14, 16, 17]. Authors in [18] examined the magnitude of the correlation between the observed variables and the factor components by manipulating sample size, number of variables, number of components, and component saturation. They concluded that the VTF ratio was important for factor stability, with more variables per factor yielding a more stable result. Authors in [19], partially confirmed this conclusion. They found that the necessary minimum indicator variables to attain factor solutions that are adequately stable relative to population factors are dependent on several aspects of any given study, including the level of communality and sample size. Similarly, authors in [20] found that when the VTF ratio increased the factor analysis solution improved.

Corresponding author: Deyab Almaleki

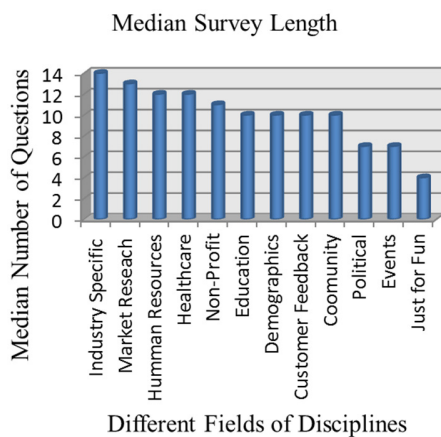


Fig. 1. Median survey length in different fields of disciplines.

The issue of variable sampling has been used extensively in conceptual development but has received almost no empirical evaluation of those that has sampled indicator variables at random from the universe of variables. The assumption of random sampling is useful to minimize sampling issues and for developing generalizability rather than a prescription for applied research procedures. Authors in [21] examined the quality of factor analytic research published between 1999 and 2009 in five leading developmental disabilities journals. They found 35% of the studies used some form of FA. However, the guidelines for using FA were largely ignored and failed to account for levels of overdetermination and communalities among measured variables. Furthermore, the authors in [19] found that there was a lack of validity in some common practice rules used in FA. Thus, anything that influences or changes variance may affect the conclusions related to FA. Researchers should determine an adequate number of indicator variables that is required to produce a stable and precise model in order to describe the latent trait. Authors in [21] investigated the effects of indicator variables on pattern recovery to determine the sufficient number of indicator variables that is likely to produce patterns that closely approximate the population pattern. They reported that the number of indicator variables can strongly affect the degree to which a sample pattern reproduces the population pattern, and that a minimum of three variables per factor is critical. The information about the adequate number of indicator variables that is required to produce a stable and precise model can be used in the design of a study and, retrospectively, in the evaluation of an existing study.

B. Adequate Sample Size

Determining sample size requirements for FA is complicated because it is dependent on other aspects of design, such as VTF and h^2 . Previous studies in FA revealed several approaches that have been used to propose guidelines for the sample size. However, most of these approaches were concerned with identifying either the Subject-To-Variable (STV) ratio or the absolute sample size, regardless of the effect of these rules on WSV. The examples below describe some reported results about sample size in the context of FA. A larger sample size is better than a smaller sample size because

it is minimizing misfit and the probability of errors. In many cases, increasing the sample size may not be possible. In medical research, it is very difficult to collect a large sample of patients suffering from a certain disease [19-21]. Investigating the minimum STV ratio or small absolute sample size to obtain the stability of the model is necessary. Only a very limited number of studies on the role of sample size in FA have investigated real or simulated small sample size. Authors in [17] investigated the minimum sample size necessary to obtain reliable factor solutions under various conditions. They concluded that under the conditions of high communality, high number of observed variables, and small number of factors, FA yields a stable estimate model for sample sizes below 50. Selecting the adequate sample size is an important decision in study design. A researcher must determine how large the sample should be and what is the most appropriate sampling frame. One problem is that the proposed recommendations vary dramatically. Clearly, the wide range in these recommendations causes them to be of rather limited value to empirical researchers. Yet, there is a need to conduct studies examining systematically the model estimate stability latent variable variance with different facets of study Experimental Design (ED) and Sampling Design (SD).

Previous research has investigated the stability of factor solutions by the examination of chi-square value (χ^2) and Overall Model Fit (OMF) indices such as Goodness-of-Fit Index (GFI), Adjusted GFI (AGFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Root Mean Square Residual (RMSR) [8, 10, 22-25]. OMF indices examined global measures of data-model fit. Examinations of Measurement Invariance (MIV) (configural, weak, and strong) were used to evaluate model stability. The effects of communality magnitude in FA have been mostly vague. Studies have revealed a varied range of communality magnitude and common practice rules [1, 3, 4, 16, 26, 27]. The communality measures the percent of variance in a given variable explained by the factors. If communalities are high, model stability in the sample data is normally very good [18, 26, 28, 29]. Authors in [20] investigated the quality of factor solutions. They found that when the communalities were high, sample size tended to have less influence on the quality of factor solutions than when communalities are low. Authors in [30] confirmed that communality magnitudes play an important role in determining the adequate sample size. Moreover, authors in [21] found that the communality magnitudes became most relevant in determining the sufficient sample size and the number of variables per component.

III. METHODS

A. Simulation Data

Simulation data are used in social science to answer a particular research question, solve a statistical problem, or improve analysis procedure techniques. Statistical program developers and research designers usually perform simulation data techniques for several reasons: gathering real data may be difficult, time-consuming, expensive, or real data sometimes violate distributional assumptions. Simulation data often lead to greater understanding of an analysis and the results one can

expect from various oddities of real-life data [31]. Simulation may approximate real-world results yet requires less time and effort and gives the researcher a chance to experiment with data under various conditions. Data can be simulated by several methods. The Monte Carlo technique is one popular method that has been used in social science since the 1940s [28]. A Monte Carlo simulation is a numerical technique that can be used to conduct experiments and repeated random sampling to simulate data for a given mathematical model. The key point of the simulation model was the development of the matrices for a 5-factor domain as an example of the factor. The statistical software package SAS was used, and the syntax code was written by the researcher.

B. Estimation Methods

Maximum Likelihood (ML) is the most common method of factor extraction that estimates population values for FA by calculating loading that maximizes the probability of sampling the observed correlation matrix from a population [12] and is often used in CFA. The current study used ML as a method of factor extraction. Authors in [21] concluded that if the data are normally distributed, ML is the best estimation option because it allows the computation of a mixed range of indexes of the goodness of fit of the model. The ML estimation method assumes that the data are independently sampled from a multivariate normal distribution with mean μ and variance-covariance matrix that takes this form: $\Sigma = LL^t + \Psi$, where L is the matrix of factor loadings and ψ is the diagonal matrix of specific variances. Authors in [32] indicated that the ML estimation method is the most precise when the data are continuous and normally distributed, but it does not provide accurate results with ordinal data or when the data violate the assumption of multivariate normality. Table I illustrates the procedure for testing model stability starting with a CFA model relative to a known factor structure for each condition involved in the study separately.

TABLE I. PROCEDURE FOR TESTING MODEL STABILITY

Test name	Symbols	Statistics guidelines
Chi-square value	χ^2	
Tucker-Lewis index	TLI	≥ 0.96 good fit
Comparative fit index	CFI	> 0.95 good fit
Root mean square error of approximation	RMSEA	0.00-0.05 very good fit 0.05-0.08 fair fit 0.08-0.10 mediocre fit

C. Procedure for Testing Stability Across Models

The procedure for testing measurement invariance was performed in order to evaluate the variation over increasingly levels of measurement invariance among models. There are different approaches that could be used to evaluate the measurement invariance among groups. The present study used a Multiple-Group Confirmatory Factor Analysis (MG-CFA) model to test invariance among the levels of STV ratios. Table II illustrates the order for testing measurement invariance starting with configural invariance (model 0). Model testing was evaluated by the chi-square difference test ($\Delta\chi^2$) between two groups [20, 30, 33], and RMSEA, CFI, and TLI were used to evaluate all the model fits. As referenced above, the criteria values suggested in [20, 34] were used in

this study: RMSEA: 0.00 - 0.05 very good fit, CFI > 0.95 good fit, and TLI ≥ 0.96 good fit. Three levels of MIV were tested.

TABLE II. PROCEDURE FOR STABILITY TESTING AMONG MODELS

M		Test name	H_0	Symbol	$\Delta\chi^2$ test	Test statistics guide
M0	Measurement Invariance (MIV)	Configural invariance	$H_0: \lambda_{group1}^1 = \lambda_{group2}^2 = \dots = \lambda_{groupg}^g$	λ : The number of factor patterns across g groups		If $\Delta\chi^2$ NS, the model shows configural factorial invariance in place
M1		Weak measurement invariance	$H_0: \lambda_j^{group1} = \lambda_j^{group2} = \dots = \lambda_j^{groupg}$	λ_j^{group1} : The factor loading of the j th indicator variable in the group	$\Delta\chi^2_{M1-M0}$	If $\Delta\chi^2$ NS, the model shows weak factorial invariance in place
M2		Strong measurement invariance	$H_0: \tau_j^{group1} = \tau_j^{group2} = \dots = \tau_j^{groupg}$	τ : The indicator variables intercept (means) of the j th indicator variable in the group	$\Delta\chi^2_{M2-M1}$	If $\Delta\chi^2$ NS, the model shows strong factorial invariance in place

Configural invariance (M0) indicates that across groups, the pattern of group1 and group2 is equivalent $\lambda_{group1}^1 = \lambda_{group2}^2 = \dots = \lambda_{groupg}^g$ where, λ represents the number of factor patterns across the gth group. Configural invariance at best indicates that the group factors are similar but gives no indication that they hold measurement equivalence. $\Delta\chi^2$ was used to judge configural invariance. For instance, if the χ^2 was not significant, the indicator variables loaded to the same factors across the groups. In other words, there were no differences in factor construct between the groups. Weak Measurement Invariance (M1) indicates that across groups, the corresponding factor loadings are equivalent: $\lambda_j^{group1} = \lambda_j^{group2} = \dots = \lambda_j^{groupg}$, where λ_j^{group1} represents the factor loading of the jth indicator variable in the group. Factor loadings represent the direct effect of the latent construct on each indicator variable, and factor loadings are represented by lambda (λ) [11]. At this level (weak measurement invariance), variables were loaded to the same factors across the group and the factor loadings across groups were equal. $\Delta\chi^2$ was used to judge weak invariance. For instance, if the χ^2 was not significant, the factor loadings across groups were equal. In other words, there were no differences in the factor loading of the indicator variables and their construct between the groups. Strong measurement invariance (M2) means that across groups, corresponding indicators' means are equivalent, $\tau_j^{group1} = \tau_j^{group2} = \dots = \tau_j^{groupg}$, where τ represents the intercept (means) of the jth indicator variable in the group. At this level (strong measurement invariance), indicator variables were loaded to the same factors across the groups. Factor loading, and indicators intercepts across groups were equal. $\Delta\chi^2_{M2-M1}$ was used to judge strong invariance. For instance, if the χ^2 was not significant, the variables intercepts across groups were equal.

D. Design

The main research question was if the stability of the simulated models varies over increasing levels of measurement invariance as a function of the following conditions and their interactions:

- Variable-To-Factor (VTF) ratio
- Subject-To-Variables (STV) ratio
- Communalities magnitude (h^2)

MGCFA was used to test the measurement invariance among the levels of STV ratios in order to evaluate model stability. Figure 2 illustrates the study design. Measurement invariance of the STV levels, as shown in Table III, were examined in each cell of the h^2 *VTF study.

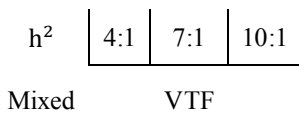


Fig. 2. Design of interaction conditions.

IV. RESULTS

Table III presents the complete findings of measurement invariance for mixed communality among levels of STV over 1000 replications where each significant p-value is marked with "*".

TABLE III. TEST OF FACTORIAL INVARIANCE FOR MIXED COMMUNALITY ACROSS VTF AND STV RATIOS AVERAGED OVER 1000 REPLICATIONS

VTF	Between groups	χ^2	M	$\Delta\chi^2$	p-value	RMSEA CFI TLI
4:1	STV=2:1 & STV=32:1	369.18	M0	---	0.0301*	0.0201 0.9893 0.9874
		---	M1-M0	---	---	---
		---	M2-M1	---	---	---
	STV=2:1 & STV=16:1	371.33	M0	---	0.0252*	0.0285 0.9791 0.9754
		---	M1-M0	---	---	---
		---	M2-M1	---	---	---
	STV=2:1 & STV=8:1	375.77	M0	---	0.0172*	0.0399 0.9602 0.9530
		---	M1-M0	---	---	---
		---	M2-M1	---	---	---

4:1	STV=2:1 & STV=4:1	386.26	M0	---	0.0065*	0.0573 0.9245 0.9104
		---	M1-M0	---	---	---
		---	M2-M1	---	---	---
	STV=4:1 & STV=32:1	340.29	M0	---	0.2084	0.0117 0.9952 0.9950
		356.07	M1-M0	15.78	0.3968	0.0117 0.9950 0.9951
		370.63	M2-M1	14.56	0.4835	0.0114 0.9950 0.9954
	STV=4:1 & STV=16:1	342.43	M0	---	0.1859	0.0164 0.9907 0.9902
		358.02	M1-M0	15.59	0.4098	0.0163 0.9904 0.9904
		372.60	M2-M1	14.58	0.4820	0.0159 0.9905 0.9910
	STV=4:1 & STV=8:1	346.89	M0	---	0.1444	0.0237 0.9825 0.9808
		362.81	M1-M0	15.92	0.3873	0.0236 0.9819 0.9810
		377.46	M2-M1	14.65	0.4769	0.0229 0.9820 0.9820
STV=8:1 & STV=32:1	329.85	M0	---	0.3402	0.0082 0.9971 0.9978	
	345.07	M1-M0	15.22	0.4356	0.0081 0.9970 0.9979	
	360.13	M2-M1	15.06	0.4471	0.0080 0.9969 0.9979	
STV=8:1 & STV=16:1	331.88	M0	---	0.3120	0.0115 0.9946 0.9957	
	346.96	M1-M0	15.08	0.4456	0.0113 0.9945 0.9958	
	362.06	M2-M1	15.1	0.4442	0.0111 0.9945 0.9960	
STV=16:1 & STV=32:1	325.34	M0	---	0.4066	0.0063 0.9980 0.9990	
	340.40	M1-M0	15.06	0.4471	0.0062 0.9980 0.9990	
	355.28	M2-M1	14.88	0.4600	0.0061 0.9980 0.9991	
7:1	STV=2:1 & STV=32:1	1257.15	M0	---	0.0006*	0.0152 0.9942 0.9938
		---	M1-M0	---	---	---

		---	M2-M1	---	---	---	---	---
STV=2:1 & STV=16:1	1262.68	M0	---	0.0004*	0.0213 0.9888 0.9879	---	---	---
	---	M1-M0	---	---	---	---	---	---
STV=2:1 & STV=8:1	1262.65	M0	---	0.0004*	0.0213 0.9888 0.9879	---	---	---
	---	M1-M0	---	---	---	---	---	---
STV=2:1 & STV=4:1	1310.82	M0	---	<0.0001*	0.0425 0.9579 0.9544	---	---	---
	---	M1-M0	---	---	---	---	---	---
STV=4:1 & STV=32:1	1170.54	M0	---	0.0686	0.0093 0.9975 0.9973	---	---	---
	1200.54	M1-M0	30	0.4656	0.0092 0.9975 0.9974	---	---	---
	1230.60	M2-M1	30.06	0.4625	0.0090 0.9975 0.9975	---	---	---
STV=4:1 & STV=16:1	1176.06	M0	---	0.0549	0.0130 0.9952 0.9949	---	---	---
	1206.12	M1-M0	30.06	0.4625	0.0128 0.9951 0.9950	---	---	---
STV=4:1 & STV=8:1	1236.47	M2-M1	30.35	0.4478	0.0127 0.9951 0.9951	---	---	---
	1191.49	M0	---	0.0279*	0.018 0.9905 0.9898	---	---	---
STV=8:1 & STV=32:1	---	M1-M0	---	---	---	---	---	---
	---	M2-M1	---	---	---	---	---	---
STV=8:1 & STV=16:1	1137.82	M0	---	0.2086	0.0062 0.9986 0.9987	---	---	---
	1167.98	M1-M0	30.16	0.4574	0.0061 0.9986 0.9987	---	---	---
	1197.93	M2-M1	29.95	0.4682	0.0061 0.9986 0.9987	---	---	---
	1143.33	M0	---	0.1773	0.0086 0.9975 0.9975	---	---	---

10:1		1173.42	M1-M0	30.09	0.4610	0.0085 0.9975 0.9976	---	---
		1203.61	M2-M1	30.19	0.4559	0.0084 0.9974 0.9976	---	---
		1122.39	M0	---	0.3133	0.0044 0.9991 0.9993	---	---
		1152.28	M1-M0	29.89	0.4712	0.0044 0.9991 0.9993	---	---
		1182.31	M2-M1	30.03	0.4641	0.0044 0.9991 0.9994	---	---
STV=2:1 & STV=32:1		2640.05	M0	---	<0.0001*	0.01241 0.9948 0.9946	---	---
		---	M1-M0	---	---	---	---	---
STV=2:1 & STV=16:1		---	M2-M1	---	---	---	---	---
		2654.33	M0	---	<0.0001*	0.0174 0.9899 0.9894	---	---
STV=2:1 & STV=8:1		---	M1-M0	---	---	---	---	---
		---	M2-M1	---	---	---	---	---
STV=2:1 & STV=4:1		2681.11	M0	---	<0.0001*	0.0244 0.9806 0.9796	---	---
		---	M1-M0	---	---	---	---	---
STV=4:1 & STV=32:1		---	M2-M1	---	---	---	---	---
		2748.81	M0	---	<0.0001*	0.0345 0.9621 0.9601	---	---
STV=4:1 & STV=16:1		---	M1-M0	---	---	---	---	---
		---	M2-M1	---	---	---	---	---
STV=4:1 & STV=8:1		2464.25	M0	---	<0.0001*	0.0076 0.9978 0.9978	---	---
		---	M1-M0	---	---	---	---	---
STV=4:1 & STV=4:1		---	M2-M1	---	---	---	---	---
		2478.53	M0	---	<0.0001*	0.0108 0.9958 0.9956	---	---
STV=4:1 & STV=16:1		---	M1-M0	---	---	---	---	---
		---	M2-M1	---	---	---	---	---

STV=4:1 & STV=8:1	2505.31	M0	---	0.0059*	0.0154 0.9918 0.9914
	---	M1-M0	---	---	---
	---	M2-M1	---	---	---
STV=8:1 & STV=32:1	2396.55	M0	---	0.1648	0.0047 0.9989 0.9990
	2441.69	M1-M0	45.14	0.4661	0.0047 0.9989 0.9990
	2486.29	M2-M1	44.6	0.4470	0.0046 0.9989 0.9990
STV=8:1 & STV=16:1	2410.83	M0	---	0.1190	0.0069 0.9980 0.9980
	2455.99	M1-M0	45.16	0.4652	0.0068 0.9980 0.9980
	2500.68	M2-M1	44.69	0.4432	0.0067 0.9980 0.9980
STV=16:1 & STV=32:1	2369.77	M0	---	0.2782	0.0034 0.9993 0.9995
	2413.98	M1-M0	44.21	0.5053	0.0033 0.9993 0.9995
	2458.99	M2-M1	45.01	0.4715	0.0033 0.9993 0.9995

VTF (4:1). The examination of the measurement invariance, begins with configural (M0) to weak (M1) and strong (M2). The findings revealed that for VTF ratios (4:1), $\chi^2_{M_0}$ showed statistically significant results when testing configural invariance: (2:1 with 32:1); (2:1 with 16:1); (2:1 with 8:1); and (2:1 with 4:1). Thus, non-invariance was established precluding further invariance testing, e.g., weak, strong, and structural. However, at higher STV ratios, e.g. groups (4:1 with 32:1); (4:1 with 16:1); (4:1 with 8:1); (8:1 with 32:1), (8:1 with 16:1); (8:1 with 32:1); and (16:1 with 32:1), $\chi^2_{M_0}$ was not statistically significant indicating configural invariance was established. Given the presence of configural invariance, testing for weak invariance was conducted. Again, chi-square difference between $\Delta\chi^2_{M1-M0}$ was not statistically significant supporting the hypothesis of weak factorial invariance between the two groups. After weak invariance was supported, the examination of the indicator intercepts was conducted. The results again supported the finding of strong invariance, e.g. the $\Delta\chi^2_{M2-M1}$ was not statistically significant. In conclusion, there were invariant factor loadings and invariant intercepts among the groups indicating that measurement invariance was achieved as described above.

VTF (7:1). The model begins with configural (M0) to weak (M1) and strong (M2). The findings revealed that for VTF ratio (7:1), $\chi^2_{M_0}$ showed statistically significant results when testing

configural invariance: (2:1 with 32:1); (2:1 with 16:1); (2:1 with 8:1); (2:1 with 4:1); and (4:1 with 8:1). Thus, non-invariance was established precluding further invariance testing, e.g. weak, strong, and structural. However, at higher STV ratios, e.g. groups (4:1 with 32:1); (4:1 with 16:1); (8:1 with 32:1), (8:1 with 16:1); (8:1 with 32:1); and (16:1 with 32:1), $\chi^2_{M_0}$ was not statistically significant indicating configural invariance was established. Given the presence of configural invariance, testing for weak invariance was conducted. Again, the chi-square difference between $\Delta\chi^2_{M1-M0}$ was not statistically significant supporting the hypothesis of weak factorial invariance between the two groups. After weak invariance was supported, the examination of the indicator intercepts was conducted. The results supported again the finding of strong invariance, e.g. the $\Delta\chi^2_{M2-M1}$ was not statistically significant. In conclusion, there were invariant factor loadings and invariant intercepts among the groups indicating that measurement invariance was achieved as described above.

VTF (10:1). The findings revealed that for VTF ratios (10:1), $\chi^2_{M_0}$ showed statistically significant results when testing configural invariance: (2:1 with 32:1); (2:1 with 16:1); (2:1 with 8:1); (2:1 with 4:1); (4:1 with 8:1); (4:1 with 16:1); and (4:1 with 32:1). Thus, non-invariance was established precluding further invariance testing. However, at higher STV ratios, e.g. groups (8:1 with 16:1); (8:1 with 32:1); and (16:1 with 32:1), $\chi^2_{M_0}$ was not statistically significant indicating configural invariance was established. Given the presence of configural invariance, testing for weak invariance was conducted. Again, chi-square difference between $\Delta\chi^2_{M1-M0}$ was not statistically significant supporting the hypothesis of weak factorial invariance between the two groups. After weak invariance was supported, examination of the indicator intercepts was conducted. The results supported again the finding of strong invariance, e.g. the $\Delta\chi^2_{M2-M1}$ was not statistically significant. In conclusion, there were invariant factor loadings and invariant intercepts among the groups indicating that measurement invariance was achieved as described above.

Table IV presents a frequency analysis of a chi-square p -values > 0.05 was tested against the null proportion $p = 0.05$ to determine if there was a statistically significant number of invariance failures. The factorial invariance result revealed that the stability of the models was varying over increasing levels of measurement as a function of VTF, STV, and their interactions.

V. DISCUSSION

The study findings refuted some of the guidelines found in the literature, e.g. authors in [18] reported that the sample size was not an important factor in determining model stability, and authors in [35] reported that the STV ratio should be no lower than 5.

The results of the current study revealed that sample size did have a strong effect on stability of the simulated models. For instance, when VTF ratio was 4:1, the mean values related to data-model fit indices were adequate at STV ratio $\geq 4:1$. However, looking at the frequency of rejections based on

conventional thresholds over the 1000 replications depicted a different conclusion. The percentage of stable (invariant) models ranged from 77% at 4:1 STV to 91% at 32:1 STV clearly indicating that larger STV ratios are related to higher stability levels with a model. These findings validated the results in [16] who reported that the percentage of invariance tests varied based on the sample size per group.

TABLE IV. χ^2 FREQUENCY AGAINST THE NULL PROPORTION OF STRUCTURAL MEAN INVARIANCE IN MIXED COMMUNALITY

VTF	Nested models	Invariance status	Frequency & percent	p-value
4:1	STV=4:1 & STV=32:1	Successful	799 79.90	<0.0001*
		Failures	201 20.10	
	STV=4:1 & STV=16:1	Successful	765 76.50	<0.0001*
		Failures	235 23.50	
	STV=4:1 & STV=8:1	Successful	712 71.20	<0.0001*
		Failures	288 28.80	
	STV=8:1 & STV=32:1	Successful	899 89.90	<0.0001*
		Failures	101 10.01	
	STV=8:1 & STV=16:1	Successful	878 87.80	<0.0001*
		Failures	122 12.20	
	STV=16:1 & STV=32:1	Successful	933 93.90	<0.0001*
		Failures	67 6.70	
7:1	STV=4:1 & STV=32:1	Successful	569 56.90	<0.0001*
		Failures	431 43.10	
	STV=4:1 & STV=16:1	Successful	513 51.30	<0.0001*
		Failures	487 48.70	
	STV=8:1 & STV=32:1	Successful	790 79.00	<0.0001*
		Failures	210 21.00	
	STV=8:1 & STV=16:1	Successful	752 75.20	<0.0001*
		Failures	248 24.80	
	STV=16:1 & STV=32:1	Successful	868 86.80	<0.0001*
		Failures	132 13.20	
10:1	STV=8:1 & STV=32:1	Successful	734 73.40	<0.0001*
		Failures	266 26.60	
	STV=8:1 & STV=16:1	Successful	670 67.00	<0.0001*
		Failures	330 33.00	
	STV=16:1 & STV=32:1	Successful	840 84.00	<0.0001*
		Failures	160 16.00	

The findings of the current study do agree with [36] where in some models an STV of 30:1 was needed to produce stable model and minimize the amount of misfit. The study findings also contradict some previous research that investigated the effect of VTF ratio on stability of factor solutions. The authors in [18] concluded that the VTF ratio was important for factor stability with more indicator variables per factor yielding more stable result.

VI. CONCLUSION

In general, this study provided an empirical evaluation of the stability of the data-model fit over increasing levels of factorial invariance for different features of design in FA. The study concluded that the stability of the models was varying over increasing levels of measurement as a function of VTF, STV, and their interactions. There were invariant factor loadings and invariant intercepts among the groups indicating that measurement invariance was achieved. For VTF ratios of 4:1, 7:1, and 10:1 the models started to show stability over the levels of measurement when the STV ratio was 4:1. Yet, the frequency of stability models over 1000 replications increased (from 77% to 91%) as STV ratio increased. The models showed more stability at or above 32:1 STV.

REFERENCES

- [1] J. S. Tanaka, "How Big Is Big Enough?: Sample Size and Goodness of Fit in Structural Equation Models with Latent Variables," *Child Development*, vol. 58, no. 1, pp. 134–146, 1987, <https://doi.org/10.2307/1130296>.
- [2] J. J. Hox, C. J. Maas, and M. J. Brinkhuis, "The effect of estimation method and sample size in multilevel structural equation modeling," *Statistica Neerlandica*, vol. 64, no. 2, pp. 157–170, 2010, <https://doi.org/10.1111/j.1467-9574.2009.00445.x>.
- [3] T. A. Holster, J. W. Lake, and W. R. Pellowe, "Measuring and predicting graded reader difficulty," *Reading in a Foreign Language*, vol. 29, no. 2, pp. 218–244, Oct. 2017.
- [4] S. Doi, M. Ito, Y. Takebayashi, K. Muramatsu, and M. Horikoshi, "Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations," *PLOS ONE*, vol. 13, no. 7, 2018, Art. no. e0199235, <https://doi.org/10.1371/journal.pone.0199235>.
- [5] T. A. Brown, *Confirmatory Factor Analysis for Applied Research, Second Edition*. New York; London: Guilford Publications, 2015.
- [6] P. J. Ferrando and U. Lorenzo-Seva, "Assessing the Quality and Appropriateness of Factor Solutions and Factor Score Estimates in Exploratory Item Factor Analysis - Pere J. Ferrando, Urbanoa, 2018," *Educational and Psychological Measurement*, vol. 78, no. 5, 2018, <https://doi.org/10.1177/0013164417719308>.
- [7] U. Lorenzo-Seva *et al.*, "Psychometric properties and factorial analysis of invariance of the Satisfaction with Life Scale (SWLS) in cancer patients," *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, vol. 28, no. 5, pp. 1255–1264, May 2019, <https://doi.org/10.1007/s11136-019-02106-y>.
- [8] D. Almaleki, "The Precision of the Overall Data-Model Fit for Different Design Features in Confirmatory Factor Analysis," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6766–6774, Feb. 2021, <https://doi.org/10.48084/etasr.4025>.
- [9] D. Almaleki, "Empirical Evaluation of Different Features of Design in Confirmatory Factor Analysis," Ph.D. dissertation, Western Michigan University, Kalamazoo, MI, USA, 2016.
- [10] O. P. John and S. Srivastava, "The Big Five Trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of personality: Theory and research, 2nd ed*, New York, NY, US: Guilford Press, 1999, pp. 102–138.

- [11] B. Thompson, "Ten commandments of structural equation modeling," in *Reading and understanding MORE multivariate statistics*, Washington, DC, US: American Psychological Association, 2000, pp. 261–283.
- [12] K. Coughlin, "An Analysis of Factor Extraction Strategies: A Comparison of the Relative Strengths of Principal Axis, Ordinary Least Squares, and Maximum Likelihood in Research Contexts that Include both Categorical and Continuous Variables," Ph.D. dissertation, University of South Florida, Tampa, FL, USA, 2013.
- [13] A. H. Blasi and M. Alsuwaiket, "Analysis of Students' Misconducts in Higher Education using Decision Tree and ANN Algorithms," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6510–6514, Dec. 2020, <https://doi.org/10.48084/etasr.3927>.
- [14] A. J. S. Morin, N. D. Myers, and S. Lee, "Modern Factor Analytic Techniques," in *Handbook of Sport Psychology*, 4th ed., M. Tenenbaum and R. C. Eklund, Eds. Hoboken New Jersey, USA: Wiley, 2020, <https://doi.org/10.1002/9781119568124.ch51>.
- [15] R. K. Henson and J. K. Roberts, "Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice," *Educational and Psychological Measurement*, vol. 66, no. 3, pp. 393–416, Jun. 2006, <https://doi.org/10.1177/0013164405282485>.
- [16] A. W. Meade and D. J. Bauer, "Power and Precision in Confirmatory Factor Analytic Tests of Measurement Invariance," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 14, no. 4, pp. 611–635, Oct. 2007, <https://doi.org/10.1080/10705510701575461>.
- [17] J. C. F. de Winter, D. Dodou, and P. A. Wieringa, "Exploratory Factor Analysis With Small Sample Sizes," *Multivariate Behavioral Research*, vol. 44, no. 2, pp. 147–181, Apr. 2009, <https://doi.org/10.1080/00273170902794206>.
- [18] E. Guadagnoli and W. F. Velicer, "Relation of sample size to the stability of component patterns," *Psychological Bulletin*, vol. 103, no. 2, pp. 265–275, 1988, <https://psycnet.apa.org/doi/10.1037/0033-2909.103.2.265>.
- [19] R. C. MacCallum and J. T. Austin, "Applications of Structural Equation Modeling in Psychological Research," *Annual Review of Psychology*, vol. 51, no. 1, pp. 201–226, 2000, <https://doi.org/10.1146/annurev.psych.51.1.201>.
- [20] K. Y. Hogarty, C. V. Hines, J. D. Kromrey, J. M. Ferron, and K. R. Mumford, "The quality of factor solutions in exploratory factor analysis: The influence of sample size, communalities, and overdetermination," *Educational and Psychological Measurement*, vol. 65, no. 2, pp. 202–226, 2005, <https://doi.org/10.1177/0013164404267287>.
- [21] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," *Psychological Methods*, vol. 4, no. 3, pp. 272–299, 1999, <https://psycnet.apa.org/doi/10.1037/1082-989X.4.3.272>.
- [22] J. C. Westland, "Lower bounds on sample size in structural equation modeling," *Electronic Commerce Research and Applications*, vol. 9, no. 6, pp. 476–487, 2010, <https://doi.org/10.1016/j.elerap.2010.07.003>.
- [23] S. Zitzmann and M. Hecht, "Going Beyond Convergence in Bayesian Estimation: Why Precision Matters Too and How to Assess It," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 26, no. 4, pp. 646–661, Jul. 2019, <https://doi.org/10.1080/10705511.2018.1545232>.
- [24] R. Jacobucci, A. M. Brandmaier, and R. A. Kievit, "A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 1, pp. 55–76, 2019, <https://doi.org/10.1177/2515245919826527>.
- [25] A. B. Costello and J. Osborne, "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis," *Practical Assessment, Research, and Evaluation*, vol. 10, 2005, Art. no. 7, <https://doi.org/10.7275/jyj1-4868>.
- [26] G. P. Brooks and G. A. Johanson, "TAP: Test Analysis Program," *Applied Psychological Measurement*, vol. 27, no. 4, pp. 303–304, Jul. 2003, <https://doi.org/10.1177/0146621603027004007>.
- [27] X. An and Y.-F. Yung, "Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It." SAS Institute Inc, 2014.
- [28] B. S. Everitt, "Multivariate analysis: The need for data, and other problems," *The British Journal of Psychiatry*, vol. 126, no. 3, pp. 237–240, 1975.
- [29] P. M. Bentler and D. G. Bonett, "Significance tests and goodness of fit in the analysis of covariance structures," *Psychological Bulletin*, vol. 88, no. 3, pp. 588–606, 1980, <https://doi.org/10.1037/0033-2909.88.3.588>.
- [30] G. D. Garson, "StatNotes: Topics in Multivariate Analysis," *North Carolina State University*. <https://faculty.chass.ncsu.edu/garson/PA765/statnote.htm> (accessed Feb. 02, 2021).
- [31] D. L. Bandalos and P. Gagné, "Simulation methods in structural equation modeling," in *Handbook of structural equation modeling*, New York, NY, US: The Guilford Press, 2012, pp. 92–108.
- [32] H. W. Marsh, K.-T. Hau, and D. Grayson, "Goodness of Fit in Structural Equation Models," in *Contemporary psychometrics: A festschrift for Roderick P. McDonald*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2005, pp. 275–340.
- [33] R. O. Mueller and G. R. Hancock, "Structural equation modeling," in *The reviewer's guide to quantitative methods in the social sciences*, New York, NY, USA: Routledge, 2018, pp. 445–456.
- [34] N. O'Rourke and L. Hatcher, *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*, 2nd ed. Cary, NC, USA: SAS Institute, 2014.
- [35] F. B. Bryant and P. R. Yarnold, "Principal-components analysis and exploratory and confirmatory factor analysis," in *Reading and understanding multivariate statistics*, Washington, DC, USA: American Psychological Association, 1995, pp. 99–136.
- [36] I. P. Cobham, "The effects of subject-to-variable ratio, measurement scale, and number of factors on the stability of the factor model," Ph.D. dissertation, ProQuest Information & Learning, USA, 1999.