10.16920/jeet/2026/v40i1/26137

Factor Structure Equivalence across Gender of a Volition Model among Engineering Undergraduates: Comparison of LRT, ΔCFI, Gamma Hat and McDonald's NCI Approaches

Chakraborty R[1], Chechi V. K[2], Gupta J[3]
[1,2,3] School of Education, Lovely Professional University, Phagwara, Punjab India
[1]rajibchakraborty07@gmail.com*, [2]chechivijay@gmail.com, [3] jyoti.bhalla@lpu.co.in

***Abstract*—Test of equal performance of a structural or measurement model in measuring a psychological phenomenon or a construct, in subjects across multiple groups, technically known as measurement invariance or measurement equivalence testing, is an essential practice in Psychometrics, which is not commonly followed in the Indian context. To promote awareness and encourage initiation of its inclusion in tool validation exercises, measurement invariance testing of the volitional model of self-regulated learning varying with respect to gender was conducted using 533 (373 males and 160 females) second and third year computer science and mechanical engineering students of the Punjab state of India. The instruments of data collection were the Academic Procrastination Scale Short Form by Yockey (2016), the Zimbardo Time Perspective Inventory Short Form by Orosz et al., (2017) and the Academic Delay of Gratification Scale by Bembenutty and Karabenick (1998). Apart from the popular approach of measurement invariance testing like Likelihood Ratio Test (LRT), lesser known invariance testing techniques like Cheung and Rensvold (2002), Steiger's Gamma's Hat and McDonald's NCI criterion which are unaffected by model complexity, sample size and overall fit measures, were administered on the collected data. The model showed acceptable overall CFA fit. Configural invariance was supported across gender. Metric invariance was rejected by LRT, ΔCFI, and Gamma Hat but supported only by McDonald's NCI by a small margin. The same result was confirmed by sensitivity analysis too. Therefore, the model should be interpreted as configurally invariant but not robustly metrically invariant across gender. Implications of the study with respect to engineering education and psychometric practices are discussed.**

Keywords*— *Measurement Invariance Testing, Measurement Equivalence Testing, Volition, Engineering Students, Engineering Education, Likelihood Ratio Test (LRT), Ordinal Omega Reliability Coefficient, Sensitivity Analysis, Cheung and Rensvold's Criterion, Steiger's Gamma's Hat, McDonald's NCI Criterion

I. INTRODUCTION

Any psychological instrument measures a latent construct which itself is composed of certain number of factors or dimensions measured with the help of their items. This structure of a construct, its related dimensions and their associated items comprise a measurement model. Every measurement model is expected to measure the construct in same way across groups or across time. This specific statistical property to be mandatorily satisfied by every measurement model of a psychological instrument is technically known as measurement invariance or measurement equivalence testing. Its definition, according to Davidov et al., (2012) is "a property of a measurement instrument, implying that the instrument measures the same concept in the same way across various sub-groups of respondents, (p.58)". The importance of this statistical property to be possessed by every psychological instrument post validation lies in the fact that the mean of a variable measured by such a tool cannot be compared across groups like male and female gender subjects or across time until it is verified as measurement invariant. The former type of measurement equivalence is called multi-group invariance and the latter type is known as longitudinal invariance. Though theoretically measurement equivalence or invariance has its mentioning way back in 1960s (Struening and Cohen,1963; Meredith, 1964), the statistical understanding was available to the researchers at the beginning of 21st century (Widaman and Reise 1997; Vandenberg and Lance, 2000).

Structural equation modeling SEM framework (Widaman and Reise, 1997) and Item response theory IRT framework (Tay et al., 2015) are the two frameworks for performing measurement equivalence, with the former, followed more commonly by practitioners. Little (1997) classified the SEM framework based measurement invariance as either belonging to category 1 or category 2.

Category 1 SEM measurement invariance testing tests the psychometric properties of the measurement scales, by conducting hierarchically constraining tests, in the order of configural invariance (Irvine, 1969; Buss and Royce 1975;

Suzuki and Rancer, 1995), metric invariance (Horn and McArdle, 1992), scalar invariance (Meredith, 1993; Steenkamp and Baumgartner, 1998) and measurement error or residual invariance (Mulle, 1995; Singh, 1995). A Category 1 test is required prior to conducting a Category 2 type, which involves studying the differences between-group differences based on their latent means, variances, and covariances.

Every psychological variable can be conceived under latent variable modeling (LVM) to be made up of certain smaller variables called its dimensions or factors, as explained by certain theory. These dimensions or factors are in turn represented and measured using items or indicators making up a psychological scale. While the construct and its dimensions / factors are latent variables existing as traits inside individuals, its items / indicators are called manifest variables since they can be seen written on a psychological scale. A graphical representation of the construct, its dimensions/factors and their respective items / indicators make up the construct's factor structure, which is population and cultural context dependent.

The psychological scale of a construct consists of a set of statements, which are the scale's items or indicators. Each of these items are the instruments or tools to measure certain factor or dimension of the construct of interest. Below every statement, certain words or phrases represent the extent to which the subjects reading the statements agree or disagree with its content and register their response by selecting any one of these words or phrases, with which a numerical score is pre-attributed indicating a gradual rise or fall in the extent of agreement or disagreement with the scale's statements by the subjects. Selection of a certain response by the subject after reading a specific statement leads to the ascribing of the corresponding score of the response to the subject for that item. Summation of the scores gathered for responding to all the items of the scale leads to the measurement of all the factors or dimensions and in turn in the measurement of the latent construct itself, in the form of the total score of the scale for a subject.

The configural invariance property possessed by a psychological scale indicates that the associated construct means the same to the subjects of both the groups like gender or batch, with construct's factor structure remaining intact (Meredith, 1993). The next level of testing is called metric invariance where whether the subjects of two groups respond to the items of the instrument in same way or not is tested, by verifying whether the factor loadings of the items are the same across subjects of different groups. When a tool is metric variant, an item of the tool may measure its associated dimension of the construct differently for boys than for girls. It can lead to measurement of the construct to be different across groups. Scales must be configural invariant to be metric invariant. The third level of testing is called scalar invariance where the intercepts of the items are tested for their stability across groups. Intercepts are the intervals of the scale with respect to its zero value. Scalar invariance is difficult to achieve, but it is a prerequisite before the means of the latent construct in both the groups can be compared. The final level of measurement equivalence testing is called the residual invariance testing where the error variance associated with each item is tested for stability across different groups. Psychological instruments are rarely residual invariant.

In essence, every psychological scale is supposed to mean the same for subjects belonging to different demographic groups like gender, batch etc. on whom the scale is administered. Subjects with same level of the construct's presence in them must understand the meaning of every item of the scale exactly the same way. Having understood the meaning of the item in the same way, these subjects must select the same response category of the item in question and hence score exactly the same as far as the measurement of the construct is concerned. Ensuring the discussed aspect of every psychological tool (which is also a measurement model) in statistical terms is called measurement invariance or equivalence.

Generally, the most common approach of measurement invariance testing, the Likelihood ratio test (LRT), is heavily dependent on chi-square test and large or small sample size of the study (Brannick, 1995; Kelloway, 1995). Also, factors like model complexity and the overall fit measures influence the results of this statistical technique. The goodness of fit estimands (properties of the data to be measured), except Root mean square error of approximation (RMSEA), tend to have smaller values with the rise in model complexity or with the increase in the number of items per factor in measurement models (Kenny and McCoach, 2003; Ruch et al., 2018; Cheung and Rensvold, 2002). The decline in the goodness of fit estimates is due to the hypothesis of confirmatory factory analysis and structural equation modelling statistical techniques to reduce small and theoretically non-significant factor loadings (strength with which an item represents its factor) to zero completely, which badly impacts the overall goodness of fit of the model under study. (Hu and Bentler, 1998; Hall et al., 1999).

Hence, the need for alternative means of estimating measurement invariance immune to the mentioned constraints was felt and addressed by (Cheung and Rensvold, 2002), when they proposed the fresh approach of testing measurement invariance of models using the difference in Comparative fit index or ΔCFI of the unconstrained and constrained models, to be less than the threshold of 0.01. The functioning of every psychological model is defined in terms of its parameters.

When these parameters are left free to vary with respect to their values, the model is said to be unconstrained. When these parameters are preset prior to the invariance testing, the model is said to be constrained. A couple of other lesser known means of measurement invariance testing similar to Cheung and Rensvold criterion and its benefits are (Steiger, 1989; McDonald, 1989).

In India, the exercise of testing and reporting the measurement invariance of psychological instruments in education sciences is not a common practice, leave alone conducting this statistical test in a manner that is free from the negative influences of model complexity, sample size and goodness of fit test estimands on the results, and hence the present study intends to add to the edifice of the literature of measurement invariance testing, procedures of the prevalent and lesser-known estimation techniques of it, through the testing of the measurement model of the construct volition (Dorrenbacher and Perels, 2015; Chakraborty and Chechi, 2020).

II. METHODOLOGY

Descriptive research design was adopted for the present study using questionnaires of the volition variables, as means of data collection from the sample subjects.

Population

Engineering students belonging to the two most sought-after branches of Engineering namely, Computer Science and Mechanical engineering AISHE (2019) in India were selected as the population of this study. The rationale for the selection of the second-year engineering students was the lack of these students' engagement in academics in general in this particular year of their study (Chasmer et al., 2015; Gardner et al., 2000), known in literature of engineering education as Sophomore slump McBurnie et al., (2012). It happens due to the raise in academic and social stressors like choice of majors (Levine and Wyckoff, 1990) experienced by these students in the second year in compared to the first year. Display of a vital self-regulated learning component, namely, volition by these students during such trying times, decide their retention or dropout from the engineering program. According to Corno and Kanfer (1993), volition is a relatively stable trait in an individual that plays an important role in making choices pertaining to a goal and working towards it in an action-oriented manner as per the action-control theory of Kuhl (1985). The need to conduct further empirical studies on this construct for gaining clarity was stressed by Zhu (2004). Bembenutty and Karabenick (20040 proposed three possible candidate variables, namely, academic delay of gratification, future time perspective and academic procrastination represent volition, thus ultimately paving way for its empirical measurement by Dorrenbacher and Perels (2015) in German context and by Chakraborty and Chechi (2020) in Indian context.

Sample

There are 94 engineering institutions (sampling frame) of the state of Punjab, are heterogeneously spread in its three regions, namely, Majha, Doaba and Malwa. The data of these institutions was obtained from the AICTE website. The region wise distribution of these institutions is shown in table I below. Also, the two engineering streams selected for the study, the class levels and gender proportion of students in these stream are also heterogeneous. To ensure representation of all these groups in the sample, stratified random sampling technique was adopted. The composition of the sample subjects consisted of four unrelated strata, namely, region, the class level, stream and gender. The final date comprised of 533 (373 males and 160 females) second and third year computer science and mechanical engineering students belonging to nine randomly selected institutions (Berry and Lindgren, 1990; Neckerud, 2013) across the three regions of the state. The homogenous stratum was represented by simple randomly selected sample subjects, all of the same average age of 19.5 years.

TABLE I
SAMPLING DESIGN

S.No.	Region-wise Total Institutions
1	Majha - 18
2	Doaba - 22
3	Malwa - 54
4	Punjab - 94

Instruments

Measurement of Future Time Perspective

The tool used to measure future time perspective in the engineering undergraduates was the Zimbardo Time Perspective Inventory Short Form developed by Orosz et al., (2017). The total scale contains 17 items which load on five factors of time perspective construct, namely, past positive with three items, past negative with four items, present hedonism with three items, present fatalistic with three items and future time perspective with four items respectively. In the present study, the items of future time perspective dimension were used. The sample item of future time perspective scale is "*I am able to resist temptations when I know that there is work to be done*." The responses of the subjects are recorded using a five-point Likert scale where 1=very uncharacteristic and 5=very characteristic. The internal consistency reliability estimation expressed through Cronbach's alpha showed that the coefficients of the five dimensions ranged from 0.68 to 0.84 and the test-retest reliabilities ranged from 0.6 to 0.76 in the original study.

Measurement of Academic Procrastination

The tool used to measure academic procrastination in the engineering undergraduates was the Academic Procrastination Scale - Short Form developed by Yockey (2016). The scale contains five items. A sample item of the scale is "*I put off projects until the last minute*". The responses of the subjects are recorded using a five-point Likert scale where 1=Disagree and 5=Agree. The internal consistency reliability coefficient expressed through Cronbach's alpha was found to be 0.87 in the original study.

Measurement of Academic Delay of Gratification

The tool used to measure academic delay of gratification in the engineering undergraduates was the Academic Delay of Gratification Scale developed by Bembenutty and Karabenick (1998). The scale contains ten items with alternative statements of instant and delayed gratifications. A sample item of the scale has the statement "*Go to a favorite concert, play or sporting event and study less for this course even though it may mean getting a lower grade on an exam you will take tomorrow, or Stay home and study to increase your chances of getting a higher grade*". The responses of the subjects are recorded using a four-point Likert scale where "1 = Definitely Choose A," 2 ="Probably Choose A," 3 "Probably Choose B," and 4 "Definitely Choose B". The internal consistency reliability coefficient expressed through Cronbach's alpha was found to be 0.77 in the original study.

Statistical Analysis

The original factor structure of the model was tested using the Structural equation modeling technique of confirmatory factor analysis CFA using SPSS AMOS Ver. 23.0. The configural and metric invariance testing were done by following the conventional Likelihood Ratio test (LRT) by checking for non-significant result of chi-square and df differences of the unconstrained and constrained structures. Since chi-square is sensitive to sample size Brannick (1995), Kelloway (1995) and Cheung and Rensvold (2002) proposed using the differences in goodness of fit index CFI of the unconstrained and the constrained structures to be less than 0.01. Their approach, estimated using SPSS AMOS Ver 23. based output, is superior to the Likelihood Ratio Test (LRT) method of measurement testing, estimated using online calculator, since it is unaffected by model complexity and sample size and not related to overall fit measures. Two more similar but less known approaches of measurement equivalence testing called the Gamma Hat by Steiger (1989) and McDonald's non-centrality index NCI by McDonald (1989) were also estimated. The former is calculated using Fan and Sivo (2007) formula (Gamma= Number of variables / (Number of variables) +(2*df*(RMSEA squared))), and the later's formula is (Mc = exp (-1/2[(TT – dfT]/N-1), where Mc is McDonald's non-centrality index NCI, TT is the chi-square value of the model, df its degree of freedom and N the sample size. These lesser-known approaches were estimated using open source excel sheets available for the same.

Procedure of Data Collection

Prior permission from the head of the engineering colleges of the state of Punjab, India was taken. This exercise was conducted by the investigator by physically visiting the selected engineering colleges and meeting the head of these institutions. They were explained the purpose of the visit and provided a formal request for data collection from the second-year computer science and mechanical engineering students of their institutions. Those institutions which provided the permission were then visited by the investigator. After explaining the purpose of the visit to the students and assuring them of confidentiality of their data and its usage solely for research purposes, their verbal consent to participate in the study was sought. Those students who provided their verbal consent were selected for the study as inclusion criteria and those who declined, were not considered for the study forming the exclusion criteria. Then, physical copies of the questionnaire were distributed to the subjects while the classes were in session and the help of the engineering faculty taking the class was sought in the smooth administration of the questionnaire and collection of the same. The students took 15 minutes of average to complete the task. Certain institutions allowed collection of the data electronically and the students of such institutions were shared the link of a google form containing the items of the variables of academic procrastination, future time perspective and academic delay of gratification. These students completed the submission of their responses in 10 minutes. The final sample size was 533.

III. RESULTS

Descriptive Statistics:

The measures of central tendency, dispersion and asymmetry of the items of volition variables were estimated using mean, standard deviation, skewness and kurtosis estimation using SPSS Statistics Ver 23.0. and are displayed below:

TABLE II
DESCRIPTIVE STATISTICS

	N	Mean	Std. Deviation	Skewness	Kurtosis

			Std.			
	Statistic	Statistic	Error	Statistic	Statistic	Statistic
ADG4	533	3.5009	.03909	.90243	-1.743	1.850
ADG8	533	3.4784	.03656	.84404	-1.485	1.142
ADG10	533	3.2026	.04421	1.02057	-.989	-.321
AP1R	533	3.0319	.06081	1.40385	-.114	-1.216
AP2R	533	2.7692	.06179	1.42653	.088	-1.359
AP3R	533	3.2627	.06118	1.41237	-.348	-1.118
AP4R	533	2.5385	.06534	1.50841	.399	-1.329
ZTP12	533	3.5497	.04661	1.07600	-.611	-.098
ZTP13	533	3.3902	.04302	.99324	-.208	-.238
ZTP14	533	3.6266	.04168	.96236	-.452	.081
Valid N (listwise)	533					

TABLE III
MEASURE OF INTER-ITEM RELATIONSHIP ESTIMATED USING PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT:

	ADG4	ADG8	ADG10	AP1R	AP2R	AP3R	AP4R	ZTP12	ZTP13	ZTP14
ADG4	1	.326	.269	-.176	-.218	-.133	-.241	.227	.169	.147
ADG8		1	.302	-.179	-.203	-.076	-.157	.149	.124	.255
ADG10			1	-.273	-.235	-.147	-.138	.183	.133	.217
AP1R				1	.312	.236	.367	-.253	-.194	-.162
AP2R					1	.246	.386	-.248	-.207	-.200
AP3R						1	.375	-.162	-.129	-.125
AP4R							1	-.329	-.114	-.225
ZTP12								1	.253	.378
ZTP13									1	.208
ZTP14										1

TABLE IV
POLYCHORIC CORRELATION BASED ORDINAL DATA RELIABILITY ANALYSIS USING PSYCH PACKAGE IN R:

S.No.	Variable	Cronbach's Alpha Coefficient	McDonald's Omega Coefficient		
1	Academic Delay of Gratification	0.56	0.57		
2	Academic Procrastination	0.65	0.67		
3	Future Time Perspective	0.54	0.56		
4	R Codes: Install (psych) Library (psych)	alpha(data_file_name)	omega(data_file_name)	a	ho)

The data obtained from questionnaire are assumed to be interval continuous data type and Pearson's Product Moment correlation is used in the estimate of internal consistency reliability Cronbach's alpha coefficient. However, this estimand assumes the condition of Tau-equivalence or equal factor loading of every item associated with its respective factor. Violation of this condition in most of the instances lead to the underestimation of the reliability coefficient of a psychological scale anywhere from 0.6 to 11 percent (Green and Yang, 2009a). A realistic estimation of reliability is done by using McDonald's Omega which is free from the underestimation of reliability coefficient under the violation of tau-equivalence condition (Hayes and Coutts, 2020; Goodboy and Martin, 2020). However, the true data type of ordinal nature obtained from the usage of Likert-scale based questionnaires still remains unaddressed. It can be resolved by estimation the reliability of the psychological scales by considering the true ordinal nature of the data and using polychoric correlation replacing Pearson's product moment correlation for reliability estimation (Gadermann et al., 2012; Zumbo et al., 2007; and Zinberg et al., (2005). A R package '*Psych*' Revelle (2019) uses simple codes to provide both the interval data based alpha and omega and the ordinal data based alpha and omega reliability coefficients of the scale.

The above table compares the estimates of the four reliability coefficients of the three variables of volition, showing the best estimation of this estimand by ordinal omega reliability coefficient estimator (the estimates marginal for being above 0.6 for all three variables) along with the underestimation of the same by the most popular Cronbach's alpha reliability coefficient (Dang and Chan, 2017).

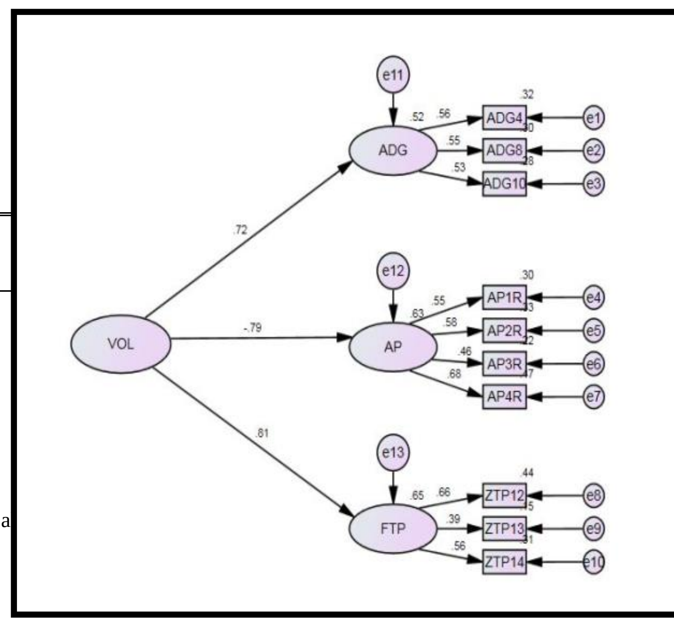

Fig. 1. Factor Loadings in the Path Diagram of the Factor Structure of Volition in Self Regulated Learning

TABLE V
FACTOR LOADINGS

S.No.	Name of the Variable	Item No	Factor Loading
1.	Future Time Perspective	ZTP12	0.66
		ZTP13	0.39

		ZTP14	0.56
2.	Academic Procrastination	AP1R AP2R AP3R AP4R	0.55 0.58 0.46 0.68
3.	Academic Delay of Gratification	ADG4 ADG8 ADG10	0.56 0.55 0.53

The path diagram of the unconstrained factor structure of volition is shown in figure 1, along with its factor loading table. It is obtained by conducted confirmatory factor analysis CFA on the hypothesized factor structure of volition with academic delay of gratification, academic procrastination and future time perspective as the component variables, verified through the gathered data. The obtained fit measures against their benchmarks (Hu and Bentler, 1995; Hu and Bentler, 1998; Thompson, 2000) are shown below:

TABLE VI
GOODNESS OF FIT ESTIMATION OF VOLITION'S FACTOR STRUCTURE

Measure	CMIN/DF	SRMR	RMSEA	CFI	TLI
Benchmark	< than 3	<0.05	0.05	0.9	0.9
Result	1.977	0.0353	0.043	0.958	0.94

Interpretation As shown in the table I, all the obtained estimands like CMIN/DF, SRMR, RMSEA, CFI and TLI are well within the desired levels of their respective benchmarks establishing the goodness of fit of the volition's factor structure. A three-factor model was specified to represent the construct volition, in which items ADG4, ADG8 and ADG10 loaded on academic delay of gratification, items AP1R, AP2R, AP3R and AP4R loaded on academic procrastination and items ZTP12, ZTP13 and ZTP14 loaded on future time perspective dimensions respectively. Cross-loadings were constrained to zero and latent factors were allowed to correlate. CFA was conducted using the conventional maximum likelihood estimator in AMOS ver. 23.0. The analysis was based on a pearson product moment correlation matrix. The latent factors were allowed to correlate, and residual error terms were assumed to be uncorrelated unless otherwise specified. No post hoc modifications based on modification indices were implemented. The model was retained in its theoretically specified form. to preserve construct validity.

	Chi-square	df	p-val	Invariant?
Overall Model				
Unconstrained	63.275	32		
Fully constrained	99.835	64		
Number of groups		2		
Difference	36.56	32	0.265	YES
Chi-square Thresholds				
90% Confidence	65.98	33		
Difference	2.71	1	0.100	
95% Confidence	67.12	33		
Difference	3.84	1	0.050	
99% Confidence	69.91	33		
Difference	6.63	1	0.010	

Fig. 2. Estimation of Configural Measurement Invariance of the Factor Structure of Volition with respect to Gender using Likelihood Ratio Test (LRT):

Interpretation: The sample consisted of 373 males and 160 females. Rather than conducting independent CFAs within each subgroup, measurement invariance analysis using multi-group CFA provided a more rigorous evaluation of whether the latent structure is equivalent across gender groups. Therefore, gender-based measurement invariance testing was conducted. The difference in the chi-square estimand of the unconstrained and the constrained factor structures is 36.56 and the difference in the degrees of freedom is 32. The p-value obtained is 0.265 which is greater than level of significance 0.05, indicating a non-significant result. It means that the factor structure is invariant with respect to boys and girls genders. Hence, the volition factor structure is configural measurement invariant and qualifies for the estimation of metric invariance.

TABLE VII
ESTIMATION OF CONFIGURAL MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING CHEUNG AND RENSVOLD'S CRITERION:

Estimates	"SRMR"	"TLI"	"CFI"	"RMSEA"	Remark	Result
Unconstrained VOL Model	0.0353	0.940	0.958	0.043	ΔCFI<= 0.01 Cheung and Rensvold Criteria (2002)	Invariant
Constrained VOL Model - Gender	0.0347	0.935	0.954	0.032		
Δ CFI	-	-	0.004	-	Yes	

Interpretation: The factor structure of volition is configural invariant with respect to gender since the difference of CFI of the unconstrained and the constrained factor structures is 0.004 desirably less than the benchmark of 0.01 of Cheung and Rensvold (2002). The factor structure hence qualifies for metric invariance estimation.

TABLE VIII
ESTIMATION OF CONFIGURAL MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING STEIGER'S GAMMA HAT TEST:

S.No.	Model	Manifest Variables	df	RMSEA	Gamma Hat	Δ Gamma Hat <=0.001	Result
1.	Unconstrained	10	32	0.043	0.99	-	
2.	Constrained Gender	10	64	0.032	0.99	0	Invariant

Interpretation: The factor structure of volition is configural invariant with respect to gender since the difference of Gamma hat of the unconstrained and the constrained factor structures is 0 which is desirably lesser than the benchmark of 0.001. The factor structure hence qualifies for metric invariance estimation.

TABLE IX
ESTIMATION OF CONFIGURAL MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING MCDONALD'S NCI TEST:

S.No.	Model	Chi-Square	df	N	NCI	Δ NCI < 0.02	Result
1	Unconstrained	63.275	32	533	0.9710	-	-
2	Constrained -Configural	99.835	64	533	0.9668	0.0042	Invariant

Interpretation: The factor structure of volition is configural invariant with respect to gender since the difference of NCI of the unconstrained and the constrained factor structures is 0.0042 which is desirably lesser than the benchmark of 0.02. The factor structure hence qualifies for metric invariance estimation.

A	B	C	D	E
	Chi-square	df	p-val	Invariant?
Overall Model				
Unconstrained	99.835	64		
Fully constrained	130.792	73		
Number of groups		2		
Difference	30.957	9	0.000	NO
Chi-square Thresholds				
90% Confidence	102.54	65		
Difference	2.71	1	0.100	
95% Confidence	103.68	65		
Difference	3.84	1	0.050	
99% Confidence	106.47	65		
Difference	6.63	1	0.010	

Fig. 3. Estimation of Metric Measurement Invariance of the Factor Structure of Volition with respect to Gender using Likelihood Ratio Test:

Interpretation: The difference in the chi-square estimand of the configural and the metric models is 30.957 and the difference in the degrees of freedom is 9. The p-value obtained is 0.000 which is lesser than level of significance 0.05, indicating a significant result. It means that the factor loadings are variant across gender. It also implies that the items of the instrument measuring volition are meant differently to both the levels of the gender, with difference in the extent to which these items measure their associated variables in the volition model. Hence, *the volition model is not metric measurement invariant* and does not qualify for further measurement invariance tests.

TABLE X
ESTIMATION OF METRIC MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING CHEUNG AND RENSVOLD'S CRITERION:

Estimates	"SRMR"	"TLI"	"CFI"	"RMSEA"	Remark	Result
Unconstrained VOL Model	0.0347	0.935	0.954	0.032	ΔCFI<= 0.01 Cheung and Rensvold Criteria (2002)	Variant
Constrained VOL Model - Gender	0.0439	0.908	0.926	0.039		
Δ CFI	-	-	0.028	-	No	

Interpretation: The volition model is not metric invariant with respect to gender since the difference of CFI of the configural and the metric constrained models is 0.028 which is undesirably higher than the benchmark of 0.01 of Cheung and Rensvold (2002). *The volition model hence does not qualify to be metric invariant* and for further measurement invariance estimations like scalar and residual error invariance tests.

TABLE XI
ESTIMATION OF METRIC MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING STEIGER'S GAMMA HAT TEST:

S.No.	Model	Manifest Variables	df	RMSEA	Gamma Hat	Δ Gamma Hat <=0.001	Result
1	Unconstrained	10	32	0.043	0.99	-	
2	Constrained - Configural	10	64	0.032	0.99	0	Invariant
3	Constrained - Metric	10	73	0.039	0.98	0.01	Variant

Interpretation: The volition model is not metric invariant with respect to gender since the difference of Gamma hat of the configural and the metric factor structures is 0.01 which is undesirably greater than the benchmark of 0.001. *The volition model hence does not qualify to be metric invariant* and for further measurement invariance estimations like scalar and residual error invariance tests.

TABLE XII
ESTIMATION OF METRIC MEASUREMENT INVARIANCE OF THE FACTOR STRUCTURE OF VOLITION WITH RESPECT TO GENDER USING MCDONALD'S NCI TEST:

S.No.	Model	Chi-Square	df	N	NCI	Δ NCI < 0.02	Result
1	Unconstrained	63.275	32	533	0.9710	-	-
2	Constrained-Configural	99.835	64	533	0.9668	0.0042	Invariant
3	Constrained-Metric	130.792	73	533	0.9471	0.0197	Invariant

Interpretation: The volition model according to McDonald's NCI test is metric invariant with respect to gender since the difference of NCI of the configural and metric constrained factor structures is 0.0197 which is desirably lesser than the benchmark of 0.02. *The factor structure hence qualifies for scalar invariance estimation by a very small margin*.

TABLE XIII
SUMMARY OF RESULTS FROM THE LIKELIHOOD RATIO TEST, CHEUNG AND RENSVOLD'S TEST, STEIGER'S GAMMA HAT TEST AND MCDONALD'S NCI TEST OF MEASUREMENT INVARIANCE TESTING OF VOLITION IN ENGINEERING STUDENTS:

S.No.	Name of the Measurement Invariance Test	Result for Configural Invariance	Result for Metric Invariance
1	Likelihood Ratio Test	Yes	No
2	Cheung and Rensvold's Test	Yes	No
3	Steiger's Gamma Hat	Yes	No
4	McDonald's NCI	Yes	Yes

Interpretation: Since three of the four tests confirm the volition model not to be metric invariant, it is summarized that the volition model in engineering students is only robustly configural invariant with respect to gender. At model level, the meaning of the construct is same for boys and girls. However, at path level, meaning of the items of the volition construct is different for these engineering subjects of two genders. The possible source of difference in understanding of the meaning of volition items among male and female subjects, can be the strong perception of gender stereotypes regarding the career choice of engineering existing among the female gender which affects their will to engage whole-heartedly with this professional education (Jimola and Jimola, 2024).

TABLE XIV
RESULTS OF THE REPEATED GENDER-BALANCED SENSITIVITY ANALYSIS FOR METRIC INVARIANCE:

Outcome	Configural Model	Metric Model	Metric-Invariance Assessment
Successful Replications	499	499	—
Median CFI	0.932	0.907	0.0642
Empirical 95% CFI range	0.902–0.957.	0.869–0.939—	—
Median RMSEA	0.0589	0.0642	—
Empirical 95% RMSEA range	0.0453–0.0707	0.0516–0.0756	—
Δ CFI criterion satisfied	—	—	8.02%
ΔRMSEA criterion satisfied	—	—	98.2%
ΔSRMR criterion satisfied	—	—	100.0%
All criteria satisfied	—	—	8.016032

Given the unequal group sizes (boys =373, girls =160), a sensitivity analysis was conducted to examine whether the conclusion of configural invariance was influenced by sample imbalance (Putnick and Bornstein, 2016). The sensitivity analysis was conducted in R version 4.5.2 using the *lavaan* package version x.x-x for multi-group confirmatory factor analysis and measurement-invariance testing. Data preparation and result summarisation were performed using *dplyr, purrr, and tibble*. Specifically, 500 random subsampling repetitions were performed, each time drawing a random sub sample of 160 boys to match the girls' group size. For each sub sample CFA was re-estimated under the configural invariance model, followed by metric invariance model. In only one, model convergence did not take place and was excluded from the fit-index summaries. The configural model demonstrated relatively stable fit across the gender-balanced subsamples, with a median CFI of 0.932 and median RMSEA of 0.0589. Imposing equality constraints on the factor loadings reduced the median CFI to 0.907 and increased the median RMSEA to 0.0642. The ΔCFI criterion was satisfied in only 8.02% of successful replications, whereas the ΔRMSEA and ΔSRMR criteria were satisfied in 98.2% and 100% of replications, respectively. These findings indicate that the conclusion of metric invariance was highly sensitive to gender-group balancing when evaluated using CFI.

IV. DISCUSSION

As a statistical technique, measurement invariance testing entered the edifice of knowledge around 1960s (Meredith, 1964). Its statistical understanding to the scientific community was accessible around 21st century (Vandenberg and Lance, 2000; Widaman and Reise, 1997). The exercise of measurement invariance testing of psychological instruments after their validation is generally recommended though conducted less by the practitioners. One of the many challenges that estimation of measurement invariance faces is the absence of a robust technique of estimation and dependence on multitudes of estimands to reach consensus on the results.

Also, the most common approach of measurement invariance testing, the Likelihood Ratio Test (LRT) method, is highly dependent on chi-square estimand and is sensitive to large sample (Brannick, 1995; Kelloway, 1995). Owing to this limitation, application of statistical techniques which are immune to model complexity, sample size and estimands of goodness of fit measures become necessary.

One such technique is the Cheung and Rensvold (2002) criterion of measurement equivalence according to which it is necessary that the difference of CFI the unconstrained and the constrained models must be less than 0.01. Other two lesser-known techniques of measurement invariance, Stegier's Gamma hat and McDonald's NCI test, make use of RMSEA and df estimands respectively. Fortunately, easy open source excel sheets are available to estimate measurement invariance of psychological instruments across groups using the mentioned techniques.

While the Volition model is found to be measurement invariant with respect to gender at the factor structure level of configural invariance (Buss and Royce, 1975; Irvine, 1969; Suzuki and Rancer, 1994) by Likelihood ratio test, Cheung and Rensvold's criterion, Stegier's Gamma hat and McDonald's NCI test, the model was found to be marginally measurement invariant at the item level through metric invariance (Horn and McArdle,, 1992) only through McDonald's NCI test.

Thus, the four tests of measurement invariance show that the boys and girls of second and third year of computer science and mechanical engineering, look at the construct of volition in a similar way (Riordan and Vandenberg, 1994). However, only McDonald's NCI test marginally displays that the items of the three variables of the model are equally good in measuring their respective latent variables in boys as well as in girls group.

Along with the novel ways of estimating measurement invariance, this study also discussed the estimation of close to accuracy ordinal omega reliability coefficient for psychological scales using the openware R-package 'psych'.

Awareness of these developments in the scientific community can make the process of psychological instruments validation more robust and scientific instead of the techniques being dependent on arbitrary rules of thumb, on which the agreement among the members of scientific community is rarely reached. The independence of estimands of these measurement invariance tests from the vital aspects of tool validation like model complexity, sample size and goodness of fit measures make these techniques way more robust and of practical use.

However, future studies must undertake refinement of items the volition model as indicated by the metric non-invariance results of this study.

V. LIMITATIONS

The items used in the study are the purified ones of the original tools for the sake parsimony (Churchill, 1979; Frohlich, 2002). Since scale purification can be sample dependent, further studies must be conducted to establish invariance of the items with respect to their functioning across the group of gender through differential item functioning (DIF) exercise. The validation of the volition tool as configural measurement invariant with respect to gender in Indian engineering students, once again establishes that the conception of volition using the three variables academic delay of gratification, academic procrastination and future time perspective by Bembenutty and Karabenick (2004) is correct at factor structure level. However, the model is not metric invariant. Hence, though the meaning of the construct volition implies same connotation in boys and girls, the items required to measure the construct equally effectively have to be constructed differently, since the items of the used tools in the present study model could not display this property. As a consequence, equality with respect to the responses of these items from their zero point also needs to be ensured which is technically known as scalar invariance, in future studies. There was an imbalance in the sample subjects in this study, with 70 per cent of the subjects being male and 30 percent being female students, as is mostly the case across India and the world (Patrick et al., 2021; Rincon et al., 2019; Jimola and Jimola, 2024; Ramachandran et al., 2020). Severe imbalance in the sample subjects with respect to gender can risk non-invariance being masked [60]. Hence, sensitive analysis was conducted which found metric invariance to be highly sensitive to gender-group balancing when evaluated using CFI. No further analysis pertaining to either conducting group-wise CFA or related to metric invariance, was found warranted thereafter. The data collection procedure involved physical distribution of the questionnaire and sharing of the google form link in electronic mode for administrative reasons of the engineering institutions which allowed the collection of the data. While the physical presence of the investigator in certain institutions during data collection ensured substantial participation of the subjects in filling the questionnaire, the same cannot be claimed in the instance of electronic mode of data collection. Finally, the study did not take into consideration the ordinal nature of the Likert-scale based data obtained from questionnaires and used maximum likelihood estimator in performing CFA and measurement invariance testing using AMOS Ver. 23.0. Future studies can incorporate the Diagonal Weighted Least Squares (DWLS/WLSMV) estimator as an appropriate alternative for performing CFA in R to address this shortcoming of the study.

CONCLUSION

The model of volition validated on engineering undergraduates demonstrated gender-level configural invariance but lacked robust support for metric invariance. Hence, the future studies can refine the items of model, conduct differential item functioning test to check for invariance in the performance of the items across the gender and adopt ordinal data based Diagonal Weighted Least Squares (DWLS/WLSMV) estimator (Osman and Yusof, 2026) in model validation and measurement invariance testing. Also, it is hoped that the Indian psychometric community adopts the lesser-known statistical techniques of measurement invariance and reliability estimation discussed in this study, in future research and contribute to validation and adaptation of psychological models which are robust enough to get replicated in multiple contexts and on multiple population.

REFERENCES

All India Survey on Higher Education AISHE 2018-19, (2019). Ministry of Human Resource Development, Department of Higher Education, Government of India.

Bembenutty, H., & Karabenick, S. A. (1998). Academic delay of gratification. Learning and Individual Differences, 10(4), 329-346. doi:10.1016/S1041-6080(99)80126-5.

Bembenutty, H., & Karabenick, S. A. (2004). Inherent association between academic delay of gratification, future time perspective, and self-regulated learning, Educational Psychology Review, 16(1), pp.35-37.

Buss, A. R., & Royce, J. R. (1975). Detecting cross-cultural commonalties and differences: Intergroup factor analysis. Psychological Bulletin, 82, 128–136.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. Journal of Organizational Behavior, 16, 201–213.3

Berry, D.A. & Lindgren, B.W. 1990. Statistics. Brooks/Cole Publishing Co., Pacific Grove, Calif.

Chasmar, J.M., Melloy, B.J. & Benson, L.B. (2015).Use of Self-Regulated Learning Strategies by Second-Year Industrial Engineering Students, Paper presented at the 122nd ASEE Annual Conference and Exposition, Seattle, WA.

Chakraborty, R. & Chechi, V.K. (2020). Network Psychometrics Based Validation of Volitional Component of Self Regulated Learning and Estimation of its Polychoric Ordinal Omega Relaibility, Mukt Shabd (UGC Care Journal), 9(6), ISSN: 2347-3150, pp: 5890-5909, doi: 09.0014.MSJ.2020.V9I6.0086781.105094.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233–255. doi:10.1207/S15328007SEM0902_5

Corno, L., & Kanfer, R. (1993). Chapter 7: The Role of Volition in Learning and Performance. Review of Research in Education, 19(1), 301–341. doi:10.3102/0091732x019001301.

Churchill, G.A. (1979), ―A paradigm for developing better measures of marketing constructs‖, Journal of Marketing Research, 16(1), pp. 64–73.

Cheung, G. W., & Rensvold, R. B. (1999). Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. Journal of Management, 25(1), 1–27. https://doi.org/10.1177/014920639902500101

Davidov, E., Dulmer, H., Schluter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. Journal of Cross-Cultural Psychology, 43, 558-575. doi:10.1177/0022022112438397

Dorrenbacher,.L. & Perels, .F. (2015). Volition completes the puzzle: Development and evaluation of an integrative trait model of self regulated learning, Frontline Learning Research, 3(4), pp:14-36, DOI: http://dx.doi.org/10.14786 /flr.v3i4.179

Deng, L. & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega, Educational and Psychological Measurements, 77(2), pp: 185-203.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. Multivariate Behavioral Research, 42(3), 509–529.

Frohlich, M.T. (2002), ―E-integration in the supply chain: barriers and performance‖, Decision Sciences, 33(4), pp. 537–556.

Gardner, J. N., Pattengale, J., & Schreiner, L. A. (2000). The sophomore year: Summary and recommendations. In L. A. Schreiner and J. Pattengale (Eds.), Visible solutions for invisible students: Helping sophomores succeed (Monograph No. 31). Columbia, SC: University of South Carolina, National Resource Center for The First-Year Experience & Students in Transition.

Gadermann, A. M., Guhn, M. & Zumbo, B.D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research & Evaluation,17(3). Available online: http://pareonline.net/getvn.asp?v=17&n=3

Green, S. B., & Yang, Y., (2009a). Commentary on coefficient alpha: A cautionary tale. Psychometrika, 74(1), pp:121-135.

Green, S. B., & Yang, Y., (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. Psychometrika, 74(1), pp: 155-167.

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, DOI: 10.1080/23808985.2020.1846135

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues and application (pp. 77-99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification, Psychological Methods, 3(4), 424-453.

Hayes, A.F. & Coutts, J.J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…, Communication Methods and Measures, 14:1, 1-24, DOI: 10.1080/19312458.2020.1718629

Horn, J. L.,& McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18, 117–144.

Hall, R.J., Snell, A.F., & Foust, M.S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. Organizational Research Methods, 2, pp:233-256.

Irvine, S. H. (1969). Contributions of ability and attainment testing in Africa to a general theory of intellect. Journal of Biosocial Science, 1, 91–102.

Jimola, A.M. & Jimola, F.E. (2024). Female students and the field of engineering: Stemming the tide of gender underrepresentation for sustainable development. International Journal of Research in Education and Science (IJRES), 10(3), 641-652. https://doi.org/10.46328/ijres.3459

Kelloway, E. K. (1995). Structural equation modelling in perspective. Journal of Organizational Behavior, 16, 215–224.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling, 10(3), 333–351.

Kuhl,.J.,(1985): "Volitional mediators of cognition-behavior consistency: Self-regulatory processes and action versus state orientation in Action control: From cognition to behavior", J. Kuhl & J. Beckmann (Ed), Handbook of physiology, (New York: Springer-Verlag) 101-128.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. Multivariate Behavioral Research, 32, 53–76.

Levine, J. & Wyckoff, J. (1990). Identification of Student Characteristics that Predict Persistence and Success in an Engineering College.

Meredith, W. (1964). Notes on factorial invariance. Psychometrika 29, 177–185. https://doi.org/10.1007/BF02289699

Meredith,W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525–543.

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. Journal of International Business Studies, 3, 573–596.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. Journal of Classification, 6, 97-103.

McBurnie, J. E., Campbell, M., & West, J. M. (2012). Avoiding the second year slump: A transition framework for students progressing through university. International Journal of Innovation in Science and Mathematics Education, 20(2), 14-24.

Nackerud, S. (2013). Collaborative Statistics Using Spreadsheets - snackeru. OpenStax CNX.

Orosz, G., Dombi, E., Toth-Kiraly, I. & Roland-Levy, C. (2017). The Less is More: The 17-item Zimbardo Time Perspective Inventory, Current Psychology, 36(1), pp: 39-47, doi: 10.1007/s12144-015-9382-2

Osman, A., & Yusof, Z.M. (2026). A hybrid factor analysis framework for ordinal data integrating CFA, ML, and DL methods. Front. Appl. Math. Stat. 12:1830237.doi: 10.3389/fams.2026.1830237

Patrick A, Riegle-Crumb C, Borrego M. (2021). Examining the Gender Gap in Engineering Professional Identification. Journal of Women Minor Science Engineeing.27(1):31-55. doi: 10.1615/jwomenminorscieneng.2020030909.

Putnick, D.L., & Bornstein, M.H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. Dev Rev. 41:71-90. doi: 10.1016/j.dr.2016.06.004.

Ruch, W., Heintz, S., Platt, T., Wagner, L., & Proyer, R.T. (2018). Broadening humor: Comic styles differentially tap into temperment, character and ability, Frontiers in Psychology, 9(6), doi: 10.3389/ fpsyg.2018.00006.

Revelle, W. (2019). Psych Package Manual, Ver. 1.9.12, Procedures for Psychological, Psychometric, and Personality Research, CRAN repository, Date of Publication: 2019-12-20, 16:40:02, UTC.

Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? Journal of Management, 20, 643–671.

Rincon, R., & Korn, R. M., & Williams, J. C. (2019, June), Examining Gender Bias in Engineering in India Paper presented at 2019 ASEE Annual Conference & Exposition , Tampa, Florida.

Ramachandran, B., Ramanathan, C. & Khabou, M. (2020). Advancement of women in engineering: past, present and future. American Society for Engineering

Education. advancement-of-women-in-engineering-pastpresent-and-future.pdf

Steiger, J. H. (1989) EzPATH: Causal Modeling. Evanston, IL:SYSTAT.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross- national consumer research. Journal of Consumer Research, 25, 78–90.

Struening, E.L., Cohen, J. (1963). Factorial Invariance and Other Psychometric Characteristics of Five Opinions About Mental Illness Factors. Educational and Psychological Measurement. 1963;23(2):289-298. doi:10.1177/001316446302300206

Suzuki, S., & Rancer, A. S. (1994). Argumentativeness and verbal aggressiveness: Testing for conceptual and measurement equivalence across cultures. Communication Monographs, 6, 256–279.

Singh, J. (1995). Measurement issues in cross-national research. Journal of International Business Studies, 26, 597–619.

Tay, L., Meade, A.W., &Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. Organizational Research Methods, 18:3–46. DOI: 10.1177/1094428114553062.

Thompson, B. (2000). Ten commandments of structural equation modeling. In US Dept of Education, Office of Special Education Programs (OSEP) Project Directors' Conference, 1998, Washington, DC, US; A previous version of this chapter was presented at the aforementioned conference and at the same annual conference held in 1999. American Psychological Association.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3, 4–69.

Widaman, KF.. & Reise, SP. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In: Bryant, KJ.; Windle, ME.; West, SG., editors. The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research. Washington, DC: American Psychological Association;. p. 281-324.

Yockey, R.D., (2016). Validation of Short Form of Academic Procrastination Scale, Psychological Reports, 118(1), pp:171-179.

Zhu, J. (2004). Understanding volition, Philosophical Psychology, 17(2), pp:247-273, DOI: 10.1080/0951508042000239066

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. Journal of Modern Applied Statistical Methods, 6, 21-29.

Zinbarg, R.E., Revelle, W., Yovel, I., & Li. W. (2005). Cronbach's Alpha, Revelle's Beta, McDonald's Omega: Their relations with each other and two alternative conceptualizations of reliability. Psychometrika. 70, 123-133. doi: 10.1007/s11336-003-0974-7.