Testing the latent structure of the ‘Jackson-5’ measure of reinforcement sensitivity across sex

By

Nadia Mercado

B.S., University of Rochester, 2015

Submitted in partial fulfillment of the requirements for the

Degree of ScM in Biostatistics in the Brown University School of Public Health

PROVIDENCE, RHODE ISLAND

MAY 2020

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions or individuals for the purpose of scholarly research.

Date______________ ____________________________________

Nadia Mercado, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date______________ ____________________________________

Nadia Mercado, Author

This thesis by Nadia Mercado is accepted in its present form by the Department of Biostatistics as satisfying the project requirements for the degree of Master of Science.

Date | Adam Sullivan, PhD, Advisor

Date | Chrystal Vergara-Lopez, PhD, Reader

Date | Christopher Schmid, PhD
Director, Master's Program in Biostatistics

Approved by the Graduate Council

Date | Andrew Campbell, PhD
Dean of the Graduate School

Preface and Acknowledgements

There are many that I would like to thank for assisting and guiding me throughout this experience. I would not have been able to go through this experience alone and I wanted to thank everyone that supported and believed in me.

To my family:

I wanted to thank my parents, Juan and Mariana Mercado for their support; you made me feel like no dream of mine was too impossible or unattainable. You allowed me to grow and supported me no matter how far I had to go. I understand that this experience has not been easy for you either and I do appreciate all the encouragement and support I have received throughout this time. Mom, thank you for the weekly phone calls and care packages. I always felt like I got to have a tiny piece of home from you. Dad, thank you for checking in once in a while and driving up to visit me and allowing me to share my life in Providence. Josue, thank you for being my brother and welcoming me to your home whenever I needed a break.

To my advisors:

I am so appreciative of how supportive and impactful a mentor can be. I feel like I have grown as a student, a professional, and as a person. Thank you for your support, Dr. Chrystal Vergara-Lopez, for believing in me and providing so many opportunities for me to learn and grow at my pace. I would also like to thank you especially for the project that would lead to this thesis. Thank you for your mentorship, Dr. Adam Sullivan, and for being the reason I decided to go to Brown University. You made me feel so secure in my decision and I am appreciative of your guidance and encouragement throughout my thesis and in this program.

To my colleagues:

I have so much to be thankful for. I never would have imagined my classmates would be such a huge part in my graduate school experience and in my life. Whenever I felt overwhelmed, there was always a classmate to assist me with learning the material or to talk to me and make me feel like I still contributed to group assignments or tell me that I would eventually learn the material, I just had to be a little more patient. I am also thankful for the memories we created outside of this program, for all the nights we shared at the graduate bar, our game nights, movie nights, every brunch, those precious girls' nights, and every other experience I was able to have with all of you. Through each of you, I have gained confidence in my ability to learn and grow my math and programming skills. I will sincerely miss everyone.

Table of Contents

List of Tables

List of Illustrations

1.0 Background

1.1 Structural Equation Modeling (SEM)

Self-reported questionnaires are often used in social and behavioral sciences to assess specific aspects of human behavior and are used to asses an underlying phenomenon which can be used to follow individuals over time or to compare different groups[1-5]. However, it is important to test the questionnaire to ensure it measures identical constructions with the same structure across different groups in order to make any meaningful comparisons. Structural equation models (SEM) are frequently used in these studies and have been applied to measure internalizing mental health problems[6], substance use[7], and other health related behaviors[8].

1.1.1 Application of Structural Equation Modeling

SEMs allow researchers to model and analyze complex causal relationships between observable variables and latent variables. On one hand, observable variables are variables that have been measured such as questionnaire items or physical traits. On the other hand, latent variables are not directly measured but are the cause for the observed variables. For example, happiness is a concept that cannot directly be measured and would be considered a latent variable. However, we may measure and observe variables such as job satisfaction, close friendships, self-assessed health and any other variable we hypothesize are indicators of the concept of happiness. These relationships are illustrated as paths in SEM diagrams.

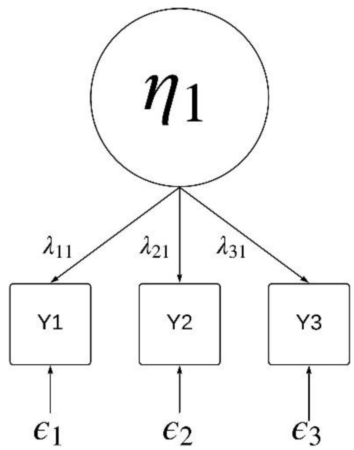


Figure 1: Path diagram of a one-factor model[9]

Note: Adapted from “Confirmatory Analysis for Applied Research” by Timothy A. Brown, 2015, Methodology in the Social Sciences, p.18, Copyright 2006 by The Guilford Press

Figure 1 is an example of a simple path diagram for a one-factor model which illustrates how observable and latent variables are modeled. The observable variables are depicted by squares (Y_1, Y_2, Y_3) and latent variables are depicted by circles (η_1). These relationships are modeled by the arrows (known as causal paths), for example, the path $\eta_1 \underset{\lambda_{11}}{\longrightarrow} Y_1$represents that η_1 has a direct effect on Y_1 (labelled λ_{11}). In addition to the direct effects ($\lambda_{11,}$ $\lambda_{21,}$ λ_{31}) there is assumed to be random error associated with the relationships (ϵ_1, ϵ_2, ϵ_3). Random errors are assumed to be uncorrelated with each other and with the latent variable. Based on the type of data and the type of structural equation modeling they can take on many different distributions, however the normal distribution is typically used.

As with any statistical model, structural equation modeling dictates that certain assumptions must hold: 1) each of the variables listed above (Y or ϵ) are normally distributed; 2) all measurement error (ϵ) is random unless otherwise specified; 3) and each path follows a linear regression.[9,10] In a traditional linear regression only the error term has to be normally distributed so SEMs require further assumptions. In addition to this, traditional statistical models have only one outcome and all relationships on the model are estimated based on that outcome. However, in a structural equation model there are many times multiple outcomes and the mixture of both observable and latent variables. Due to these complex relationships different software and algorithms are required than more traditional models.

Several programming applications that can be used to implement the various models of Structural Equation Modeling such as: LISREL[11], SAS/CALIS[12], STATA[13], and Mplus[9,14]. Mplus is rapidly becoming more ubiquitous in Structural Equation Modeling throughout psychology and social sciences research as it has relatively simple commands to develop and handle complex models and data[9]. In addition to the ability to handle complex models and data, Mplus is updated often and has a wide user base making it the choice of many SEM users.

1.1.2 Confirmatory Factor Analysis (CFA)

There are many models in which SEMs are used. One such model is for confirmatory factor analysis (CFA). CFA is an analytical method commonly used in psychometric studies which allows the researcher to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists[15,16]. The observed variables are known as indicators which can be comprised of questionnaire items, behavioral observations, or test scores; essentially this can include any measurable construct the research question requires. The latent construct, or factor, is a theoretical idea which cannot directly be observed. Instead the latent construct is manifested in the observed indicators and is the believed to be the cause of the indicators we can observe in the real world.

The items or indicators load onto the latent construct y, a matrix containing the items:[17]

$$y = \Lambda\eta + \epsilon$$

$$\Lambda = mxp\ matrix\ containing\ the\ factor\ loadings$$

$$\eta = px1\ latent\ variable\ vector$$

$$\epsilon = \text{mx1 matrix of errors}$$

There is also a correlation matrix where we assume correlated factors where Ψ is a pxp correlation matrix of p pindicators:[9,17]

$$\Sigma = \Lambda\Psi\Lambda' + \Theta$$

$$\Sigma = pxp\ correlation\ matrix\ of\ p\ indicators$$

$$\Lambda = pxm\ matrix\ containing\ the\ factor\ loadings$$

$$\Psi = mxm\ latent\ variable\ covariance\ matrix$$

$$\Theta = pxp\ covariance\ matrix\ of\ errors$$

Although there is no agreed upon cutoff, for each list of factors that are considered to be part of a single latent construct, the general convention is that each of the loadings for each factor are considered substantial if the standardized loadings were greater than 0.3[18]. CFA has become an important analytical tool to estimate the scale reliability of various psychometric evaluations used in the social and behavioral sciences[9]. The true value of CFA is in reducing the number of observable variables that relate to the latent factors. This can make the model more interpretable as well as reduce the chance of multicollinearity in the model.

1.1.3 Model Fit Statistics

Traditionally, a common method for assessing model fit used a χ^2 statistic. However, it is not typically advised to measure model fit by χ^2 as this can be artificially inflated by sample size or for small N, the sample may not follow a χ^2 distribution[17,19]. We can adjust for this effect by measuring fit by the CFI, where the CFI can be measured as follows: [9,18]

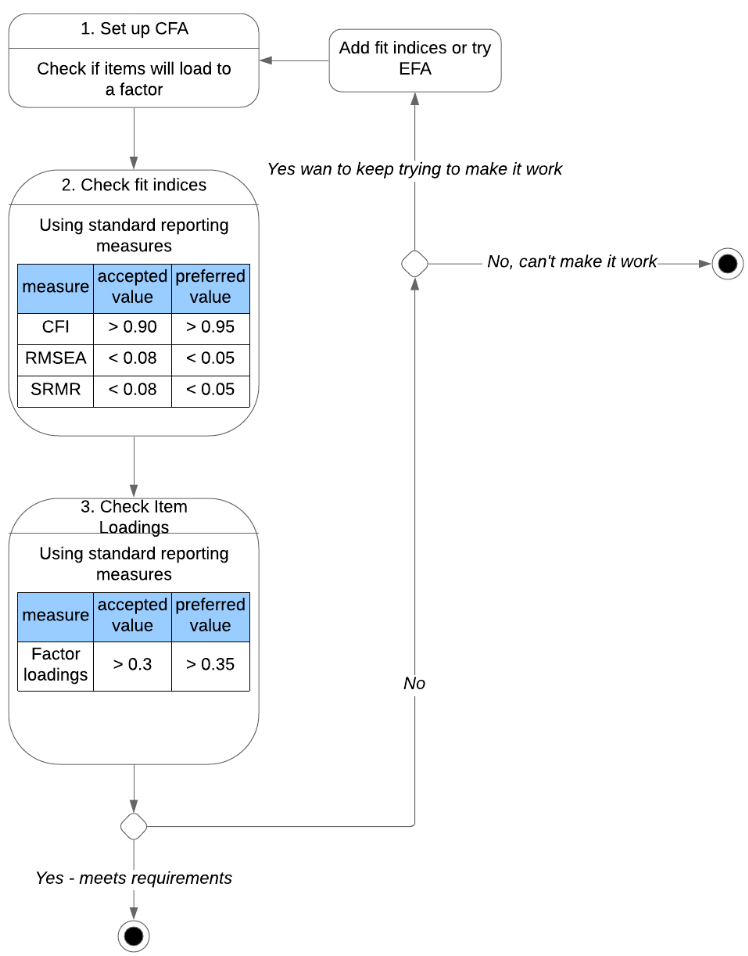
$$CFI = 1 - \frac{max[\chi_T^2 - df_T, 0]}{max[(\chi_T^2 - df_T), (\chi_B^2 - df_B), 0\,]}$$

where *T* denotes the target model and *B* denotes the baseline or null model. In practice, a CFI > 0.90 indicates acceptable fit, while a CFI > 0.95 is preferred[17,19]. Other measures which can evaluate closeness of fit include root mean square error of approximation (RMSEA) (values < 0.08 indicate good fit, <0.05 preferred) and standardized root mean square error residual (SRMR) (values < 0.08 indicate good fit, <0.05 preferred)[17].

It is important to not only look at the factor loadings but also the fit statistics to confirm if the factors measured load onto a known latent construct. If the model fit is poor the CFA model can be revised to improve the fit of the model. A source of misspecification could be from the correlated errors (Ex. ϵ_1- ϵ_3 in Figure 1). Unless specified, the model will assume that all the covariation among the items that load onto a factor is due to that latent dimension and that all measurement error is random[9].

In some cases, this error could also be partially shared variance for items within the same latent factor or a source outside of a common latent factor. This may be justified in cases where the items appear similar in wording or susceptibility to other influences[9]. In Mplus this is known as adding fit indices in which the error terms from the indicators are allowed to be correlated in the same or different latent factor in the model. This can help improve the model because while each of the items should have been designed to measure one part of the latent construct and be independent of the other items, this may not always be the case. By adding these measures, it could help explain the model better and it also accounts for more error in the model as a whole which may improve model fit. If this still does not improve model fit, it may be worth deconstructing the latent factor and performing exploratory factor analysis (EFA) to reconstruct the latent construct as the items do not represent the latent construct.[18] Figure 2 provides an overview on the process involved when performing CFA.

Figure 2: Schematic overview of confirmatory factor analysis (CFA)



1.1.4 Measurement Invariance (MI)

Measurement Invariance is an analytical technique that can be used to demonstrate that the participants across different groups or across different time points do interpret the individual questions in the same way, thereby allowing researchers to make meaningful comparison across groups [16,20,21]. If the test was created to be administered in a heterogenous population and the assumption of measurement invariance cannot be established, then the test may be biased against specific groups such as by sex or culture[9]. These tests are similar to a t-test or ANOVA which can be used to compare observed group means, however, some researchers may prefer an SEM approach as these comparisons can be made in the context of a latent variable measurable model[9]. The framework from CFA lends itself to evaluate the equivalence of measurement models among distinct groups. A multiple-group CFA uses a system of simultaneous CFA analysis for each group or timepoint. Each group will be evaluated by the latent structure individually and assessed for equivalence of the measurement and structural solution[9]. There are three critical steps that should be considered: configural invariance (equivalence of model form – do the constructs have the same pattern of free and fixed loadings across groups); metric invariance (equivalence of factor loadings – each item contributes to the latent construct to a similar degree across groups); scalar invariance (equivalence of item intercepts or thresholds – mean differences in the latent construct capture all mean differences in the shared variance of the items)[15,16].

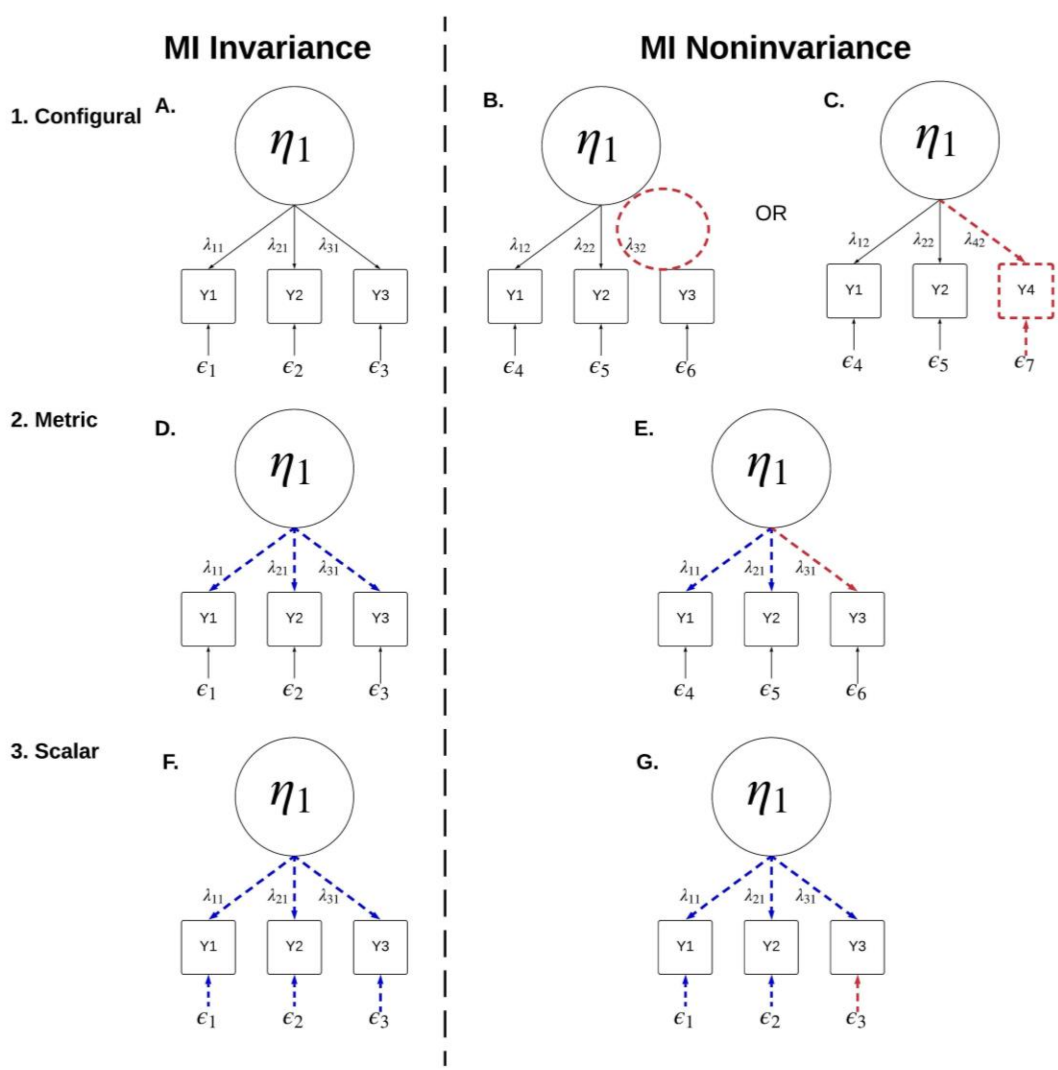


Figure 3: Path diagram illustrating Measurement Invariance and Noninvariance

Note: In measurement invariance tests, all models are fit to each of the groups (ex. male and female) and parameters are constrained to be equal across groups. For the invariance models illustrated in D and F, the bolded paths in blue are the are set to be equal across groups. In the noninvariance groups illustrated in B, C, E, and G, the bolded paths in red are examples where when this is applied to one group results in measurement noninvariance for that respective step (configural, metric, or scalar invariance).

Configural invariance is the baseline model which is used to evaluate to which degree the different levels of MI are supported by the data. In this model, the values between each factor loading or intercepts are not constrained between the groups or timepoints that are being compared in the study[18]. Configural invariance tests for basic model fit information, and if the values in the factor loadings are roughly equivalent and in the same direction across groups[18]. Specifically, we would like to confirm in configural invariance if the latent factor (η1) is represented by the same items (Y_1-Y_3) and has roughly the same patterns of factor loadings (λ_{11}- λ_{31} and λ_{12}-λ_{32}) across groups. This means we have a diagram for Figure 3A for each group (ex. one for males and one for females). If we find configural noninvariance, could mean that the pattern of factor loadings differs across the groups such as one item is only represented by a latent factor in one group (Figure 3B) or a different item loads onto a latent factor (Figure 3C). For example, this means we could have one group (ex. male) be represented by Figure 3A and another group (ex. female) could be represented by either Figure 3B or 3C. If the model meets configural invariance, metric invariance can be evaluated [15].

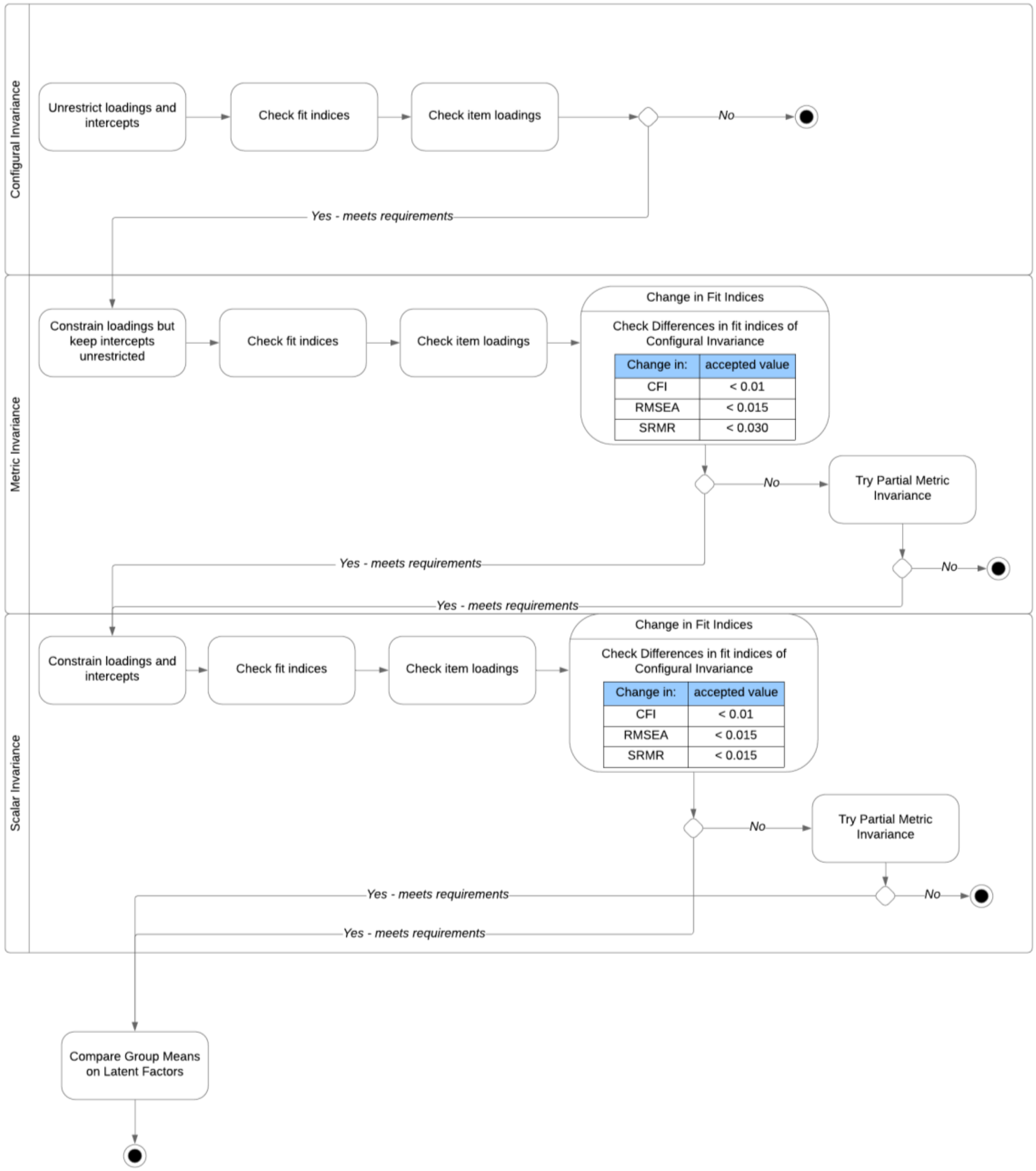
Metric invariance determines if the latent constructs are comparable at the latent level (i.e. across groups or time by the factor variance) [15,16]. Here, the values between each factor loading are constrained a reference group and the item intercepts are still freely estimated. In metric invariance, we are confirming if we constrain the loadings of the items on the latent factor to be equivalent across groups (represented by the blue arrows in Figure 3D and 3E) would significantly change the fit of the model. As previously described, if we find metric invariance we would have a diagram as illustrated in Figure 3D for each of the groups. If we find metric noninvariance, this could indicate that at least one loading is not equivalent across groups (represented by the red arrow in Figure 3E). If the model meets metric invariance, scalar invariance can be assessed[15].

Scalar invariance determines if the latent constructs are comparable at the manifest level (i.e. across all groups and time by the observed means). In this model, the factor loadings and the item intercepts are constrained to a reference group. For scalar invariance, we are confirming if we constrain the loadings and item intercepts of a latent factor across groups (represented by the blue arrows in Figure 3F and 3G) would significantly change the fit of the model. If we are able to confirm scalar invariance, we would be able to represent each group by the diagram in Figure 3F. If we find scalar noninvariance, this may indicate that at least one item intercept differs across groups (represented by the red arrow in Figure 3G)[16]. If at least partial scalar invariance is established, then the mean differences in each of the groups can be compared at the latent level and generally indicates that there are no differences across how the groups responded to the items for that particular latent factor[22]. In summary, if we have measurement invariance at each stage, this means that across groups (ex. sex, culture, time) we should have one diagram from Figure 3A, 3D, and 3F for each subgroup.

For each model, it is critical to compare the model fit information for each. While attempts have been made to try to impose a variety of standards for comparing fit indices within and between these models[15,16,21], some have argued that criteria to evaluate fit indices need to vary by the specific model and research question[16]. The fit indices that are typically used as criteria for evaluation are the same as previously cited for CFA: χ^2, degrees of freedom (DF), Comparative fit index (CFI), Root Means Square Area of Approximation (RMSEA), and the root mean square error of approximation (SRMR). In Configural, Metric, and Scalar Invariance, the fit indices should follow the same requirements as previously mentioned for CFA.

In addition, models would have been considered to meet MI need to have a nonsignificant change between each subsequent model that becomes more constrained at each step. In metric and scalar invariance, we test for measurement invariance by evaluating the change in the Comparative Fit Index (CFI) where a change in CFI less than 0.1 suggests MI and a change greater than 0.1 suggests a systematic bias in measurement[16]. While prior recommendations have evaluated MI using chi-square difference values, they have been found to be highly sensitive to differences in large samples and less sensitive to non-invariance [19,22]. Therefore, the present study utilizes changes in CFI to test for MI which have been found to be less sensitive in larger sample sizes [23]. Other fit indices such as χ^2, degrees of freedom (DF), Root Mean Square Area of Approximation (RMSEA), and the root mean square error of approximation (SRMR) will also be reported as they have also been recommended to test for MI[16]. If MI cannot be established and systematic bias in measurement is detected, then partial invariance models could be evaluated by releasing parameters that vary statistically significantly across groups by freely estimated those parameters instead of constraining them across groups and examined the statistical fit and evaluate the MI [20,24,25]. Figure 4 provides an overview of the iterative process for MI.

Figure 4: Schematic overview of Measurement Invariance (MI)



For more details and an explanation on how to code CFA in MPlus, see Appendix A. The code has mostly been adapted from Hoffman's lecture on multiple group CFA analysis[26].

1.2 'Jackson-5' scale

The original Reinforcement Sensitivity Theory (o-RST) was a scale developed by Jeffrey A. Gray and McNaugthon[27-29] which theorizes that human behavior is guided by individual differences in reactivity to reward and punishment. This is measured by three systems: Behavioral Activation System (BAS), Behavioral Inhibition System (BIS), and the Fight-Flight System (FFS)[28]. Research in psychology and the social sciences support the notion that these scales are useful for investigating the development of psychopathology, substance use, and other health related behaviors[6,30,31].

Despite being one of the more recognized approaches to measure these different biological systems, some of the criticisms of the o-RST argue that while Gray postulated that the three biological systems were separate and independent, many have argued that these systems may interact with each other. Though other iterations of the RST have been developed, they have also been unable to create a scale with anatomical independence between the three systems [29,32]. Because of this, researchers have been unable to use the scale to measure differences in groups for certain psychopathologies such as anxiety or depression[33-35]. While there are several iterations of the o-RST[29,34,36], the 'Jackson-5' r-RST was created to address many of these criticisms and has been of the most frequently used measures[37,38].However, the revised scale also introduced sex as a parameter in the structural model which may suggest that the functioning of the questionnaire may differ across sex. To date, there are only two studies that have tested other measures of reinforcement sensitivity across sex[39,40]. Although there are a number of publications that have risen in support of the 'Jackson-5' scale [32,33,35,38,41], there is lack of literature which prompts the question if the psychometric properties of this measure are equivalent across sex.

2.0 Statement of Purpose:

Assessing a psychometric scale's ability to operate similarly across groups is critical as violating assumptions of equivalent functioning across groups can result in invalid causal inferences[42]. Currently, only two articles have been found that test the measurement invariance of the reinforcement sensitivity scales across sex and used the o-RST scales[39,40], therefore, it is important to test the r-RST to see if assumption of equal instrument functioning across sex is met. The present study interviewed 185 undergraduate students in a public university in the Northwestern United States and were administered the 30-item 'Jackson-5' r-RST questionnaire [30]. This study is a secondary analysis of the original study and will be using CFA and MI analysis to test the factor structure of the 'Jackson-5' measure across sex. Lastly, the analysis will also look at techniques outside of structural equation modeling and test to see if similar or different results arise from alternative methods to test the 'Jackson-5' scales such as linear regression.

3.0 Methods

3.1 Participants

A total of 185 participants completed a series of self-reported questionnaires, which included the ‘Jackson-5’ scale, under the direct supervision of a trained research assistant and overseen by a licensed clinical psychologist. Participants were enrolled in an undergraduate level psychology course at a public university in the Northeastern region of the United States with an average age of 19 (SD = 1.42); approximately 44% male and 56% female; and identified as Caucasian (72%), Hispanic (8%), African American (6.5%), other (6.5%), Asian (6%), and American Indian (1%). Participants that successfully completed the self-reported questionnaire received course credit in partial fulfillment of a course requirement. All participants gave consent to take part in the study through an IRB-approved protocol.

3.2 Measures

3.2.1 Depression Diagnosis

The Structured Clinical Interview for the DSM-IV [43] was utilized to asses both mood disorder and psychosis. The SCID is a widely administered[44-47], clinical assessment that uses diagnostic criteria from the DSM-IV. Administering of the SCID was conducted through semi-structured interviews, which were recorded in full to ensure procedural consistency and assess inter-rate reliability (*Fleiss K*= 0.92).

3.2.2 *Revised Reinforcement Sensitivity Theory*

The Jackson-5 scales[37] were used to measure revised Reinforcement Sensitivity Theory (r-RST). This scale is derived from Gray’s original RST [27,28], the Jackson-5 scales are a brief measurement of the five behavioral facets of revised Reward Sensitivity Theory: BAS, BIS, and FFFS (fight, flight, and freeze). In total, 30 questions are administered (six items per construct) to assess the five different systems involved in reward and punishment sensitivity. Items were scored on a scale of 1 (completely disagree) to 5 (completely agree). See Table 1 for questionnaire items[37].

4.0 Approach to analysis

4.1 Structural Equation Modeling

All latent structure analyses were conducted using Mplus (version 7.0)[14] and results were compiled from R using the MPlusAutomation package [48] in R (version 3.5.0). Using confirmatory factor analysis (CFA), we tested the latent structure of the 'Jackson-5' measure of reinforcement sensitivity for each of the five behavioral factors (BAS, BIS, Fight, Flight, Freeze) without adjusting for sex.

We then considered measurement invariance to test the latent structure of the 'Jackson-5' measure across sex. We applied a series of nested model tests within a multigroup CFA framework. The first model tested for configural invariance (no constraints across sex) and verified if each of the items displayed a similar pattern of factor loadings across sex. The second model tested for metric invariance (constraining factor loadings to be equal across sex) and checked the statistical fit of each model relative to the configural invariance model. Lastly, scalar invariance was tested (constraining factor loadings and item intercepts to be equal across sex) and comparing the fit of this model relative to the metric invariance model.

In the event that either metric or scalar invariance was rejected (i.e. the fit of the nested model varied significantly from the previous model), we tested for partial measurement invariance. This is accomplished by releasing parameters that have been constrained by measurement invariance which have been identified as noninvariant and then retesting the fit of the partial invariant model to the previous iterative model, according to current recommendations in literature[16].

4.2 Linear Regression

We also considered simple linear regression models to evaluate the relationships within the structural model of the Jackson-5. We also considered mediation analysis as well to avoid some of the strong assumptions in the SEM model in Jackson's paper. This analysis was conducted in R (version 3.5.0).

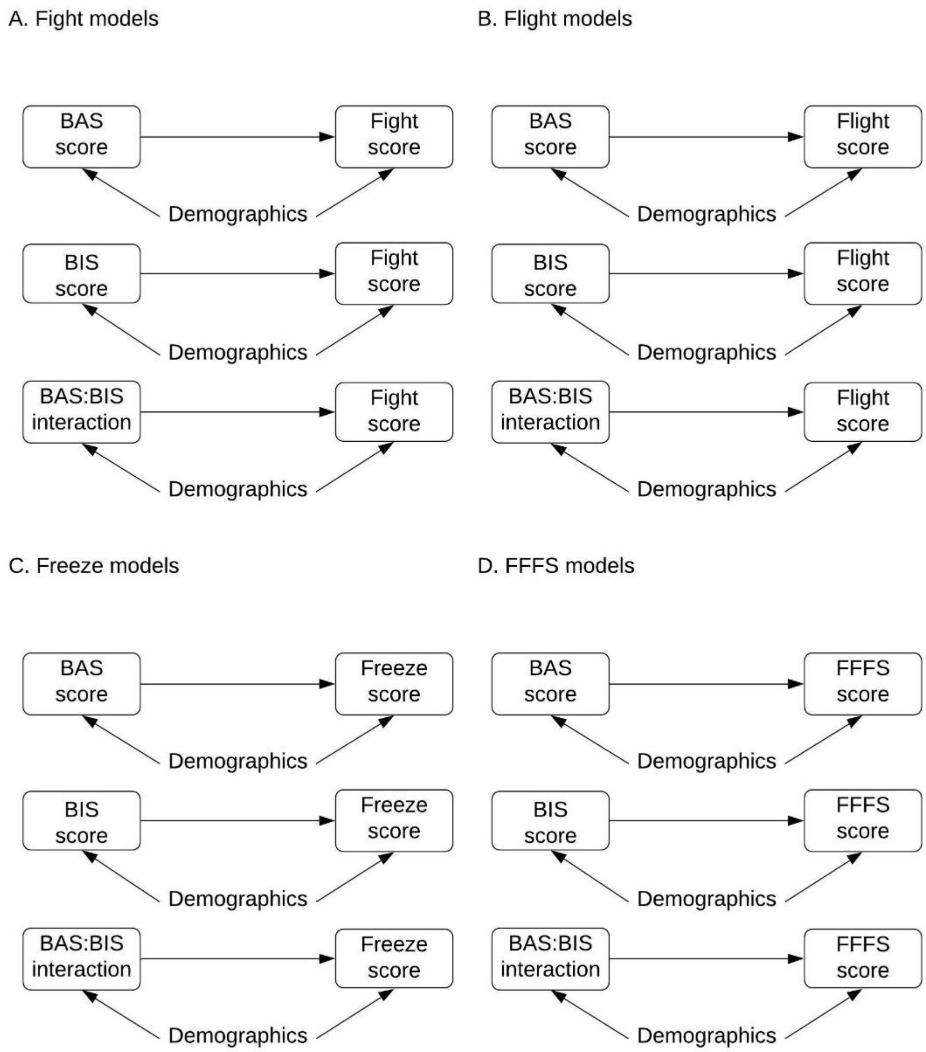
For this analysis, we collapsed each of the items into a single variable to represent each of the factors we are studying: BAS, BIS, Fight, Flight, Freeze, FFFS, Anxiety, and Depression. We accomplished this by converting the responses into quantiles to scale each of the items and sum the items together to create each factor. In order to create the FFFS, the Fight, Flight, and Freeze items were once again transformed by the quantiles and summed together. This is similar to the standardization which occurs for each of the items in CFA [18]. Other covariates that were adjusted for in the regression models included age and sex.

We first tested to see if the response of each individual behavioral trait in the FFFS model (Fight, Flight, and Freeze) would produce different results (such as different coefficients, by magnitude or significance) with either BAS and BIS modeled separately and together in the same model adjusting for age and sex.

As Fight, Flight, and Freeze were developed with the convention that they all measure behavioral avoidance traits[37], within each of the regression models that are being tested, we would expect that there should be no significant difference within the magnitude and significance of each of the covariates – particularly with BAS and BIS.

If the direction, magnitude, or significance of the coefficients are different, it may indicate that the FFFS as currently structured is not an appropriate factor and future studies may need to revisit the at the item-level how the FFFS should be constructed. A schematic of the regression analysis is available in Figure 5.

Figure 5: Schematic of Regression Analysis



5.0 Results

5.1 Results – Structural Equation Modeling

5.1.1 Correlations and descriptive statistics

Table 2 describes the means, standard deviations, and correlations for all questionnaire items for the ‘Jackson-5’ scale. The items on the ‘Jackson-5’ scale ranged in skew between -1.67 and 1.19 (mean=-0.27; SD=0.73); kurtosis ranged from 1.78 to 7.43 (mean=2.47; SD=1.40).

Table 1 – Standardized factor loadings across entire sample and by sex

	Collapsed across sex	Estimated for females	Estimated for males
BAS			
1. I like to do things which are new and different.	.67**	.64**	.70**
2. I like to do things spontaneously.	.58**	.67**	.45**
3. I actively look for new experiences.	.85**	.85**	.86**
4. I have a feel for how things work.	.43**	.46**	.42**
5. I look for new sensations.	.77**	.75**	.83**
6. I am excited about what is new in my field.	.48**	.55**	.36**
BIS			
1. I aim to do better than my peers.	.27**	.28**	.22
2. I want to do well compared to my peers.	.35**	.30**	.43**
3. I like my peers to know I am doing well.	.72**	.72**	.72**
4. I prefer to work on projects where I can prove my ability to others.	.73**	.86**	.51**
5. I want to avoid looking bad.	.20**	.20	.23
6. I avoid work that makes me look bad.	.26**	.31**	.21
Fight			
1. I would fight back if someone hits me first.	.75**	.77**	.59**
2. When provoked, I easily get into a fight.	.61**	.64**	.61**
3. If a burglar broke into my house, I would immediately look for a weapon.	.66**	.59**	.63**
4. If I caught somebody stealing my belongings, I would attack.	.78**	.79**	.72**
5. If I think somebody is going to hit me, I hit them first.	.65**	.65**	.56**
6. If somebody does something bad to me, I would retaliate.	.67**	.57**	.72**
Flight			
1. If I am approached by a suspicious stranger, I run away.	.28**	.10	.42**
2. I am likely to run if harassed by a stranger in an unfamiliar place.	.33**	.27*	.27*
3. If a dog barks at me, I would run away.	.63**	.68**	.56**
4. If the fire alarm rings, I immediately rush out of the building.	.50**	.45**	.52**
5. I can’t help but feel terrified if I see a dangerous animal.	.80**	.80**	.78**
6. I used to hide behind a chair as a child when I was watching a frightening TV show.	.19*	.18	.11
Freeze			
1. If something very bad was just about to happen to me, I would just stop.	.32**	.33**	.14
2. If I got scared in my bed at night, I would remain motionless.	.42**	.29*	.54**
3. I don’t know what to say if a stranger is rude to me in the street.	.63**	.62**	.71**
4. If my box told me to do two contradictory things, I would not know what to do.	.62**	.72**	.47**
5. If there is a choice of products in the shop, I find it hard to decide what to buy.	.30**	.34**	.25
6. In a crowd, my mind freezes and I never know what to say.	.46**	.52**	.35*

Note: Factor loadings estimated separately for females and males via a multi-group confirmatory factor analysis (no parameter estimates were constrained across sex);
** = p< .01; * = p<.05

Table 2 - Pearson Correlations, Means, and Standard Deviations of ‘Jackson-5’ items

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.	M	SD
1. BAS1	1.00																														4.28	0.76
2. BAS2	.59**	1.00																													4.05	0.90
3. BAS3	.59**	.52**	1.00																												3.91	0.95
4. BAS4	.25**	.19**	.31**	1.00																											4.09	0.73
5. BAS5	.51**	.43**	.65**	.39**	1.00																										3.79	0.95
6. BAS6	.28**	.22**	.40**	.41**	.35**	1.00																									4.12	0.75
7. BIS1	.02	-.14	.10	.05	.02	.22**	1.00																								4.19	0.85
8. BIS2	.09	-.10	.01	-.02	.05	.10	.70**	1.00																							4.51	0.67
9. BIS3	.00	.04	.15*	.15*	.19**	.22**	.17*	.23**	1.00																						3.50	1.10
10. BIS4	.01	.04	.04	.13	.11	.26**	.22**	.26**	.53**	1.00																					3.68	1.01
11. BIS5	.01	-.07	-.01	-.14	-.05	.07	.14	.22**	.14	.11	1.00																				4.25	0.79
12. BIS6	-.12	-.03	-.10	-.05	-.10	-.01	.03	.12	.21**	.17*	.47**	1.00																			3.32	1.17
13. Fight1	.12	.20**	.08	.12	.20**	.02	.05	.07	.05	.02	.03	-.05	1.00																		3.82	1.37
14. Fight2	-.07	.05	.01	-.01	.04	-.08	.00	.07	.13	.08	-.02	.20**	.48**	1.00																	2.41	1.22
15. Fight3	-.01	-.03	.05	.15*	.12	-.07	.01	.07	.24**	.20**	.04	.05	.53**	.35**	1.00																3.80	1.29
16. Fight4	-.01	.10	.05	.05	.12	.01	-.00	.02	.11	.06	-.06	.03	.58**	.46**	.54**	1.00															3.35	1.18
17. Fight5	.05	.07	.02	-.02	.12	-.15*	-.08	-.02	.13	.07	-.05	.11	.43**	.40**	.41**	.55**	1.00														1.93	1.01
18. Fight6	-.10	-.01	-.10	-.00	-.01	-.19*	-.03	.02	.06	-.02	-.08	.05	.51**	.46**	.42**	.49**	.48**	1.00													3.15	1.19
19. Flight1	-.11	-.05	.07	.01	-.04	.07	.19**	.14	.17*	.12	.14	.23**	-.17*	.02	-.15*	-.03	-.07	-.17*	1.00												2.96	1.08
20. Flight2	.04	.00	.08	-.04	-.02	.01	.08	.10	.08	-.02	.12	.20**	-.27**	-.09	-.14	-.02	-.06	-.16*	.71**	1.00											3.43	1.19
21. Flight3	-.16*	-.05	-.04	-.09	-.10	-.07	-.10	-.20**	.02	-.05	-.11	.01	-.03	-.01	-.07	-.07	.12	.06	.13	.14	1.00										2.18	1.13
22. Flight4	-.13	-.16*	-.03	-.21**	-.11	-.02	.18*	.08	-.04	-.11	.02	.02	-.15*	.00	-.09	-.03	-.11	-.13	.23**	.23**	.31**	1.00									2.58	1.17
23. Flight5	-.07	-.06	.03	-.10	-.09	-.09	.06	.01	-.04	-.03	.09	.11	-.09	-.05	-.11	-.08	.03	-.05	.21**	.27**	.51**	.38**	1.00								2.54	1.19
24. Flight6	-.04	-.05	-.01	-.06	-.04	.00	-.03	-.04	.09	-.03	.02	.03	-.11	-.01	-.04	-.10	-.06	-.17*	.09	.08	.08	.19**	.13	1.00							1.88	1.10
25. Freeze1	-.06	-.06	.04	-.08	-.04	-.01	.03	-.02	.16*	.00	.05	.13	-.14	-.08	-.13	-.16*	-.12	-.09	.22**	.26**	.24**	.29**	.30**	.27**	1.00						3.08	1.19
26. Freeze2	-.13	-.10	-.11	-.10	-.13	-.01	.01	.03	.01	.02	.12	.17*	-.16*	.04	-.05	.00	-.14	-.06	.14	.12	.13	.06	.21**	.09	.21**	1.00					3.30	1.27
27. Freeze3	-.14	-.13	-.17*	-.32**	-.24**	-.12	.02	.05	.02	-.04	.26**	.17*	-.34**	-.16*	-.25**	-.21**	-.22**	-.28**	.23**	.22**	.02	.21**	.23**	.14	.17*	.30**	1.00				2.66	1.22
28. Freeze4	-.09	-.08	-.09	-.21**	-.20**	-.09	-.00	.04	.04	-.01	.11	.13	-.20**	-.03	-.06	-.09	-.17*	-.19*	.15*	.16*	-.02	.27**	.09	.10	.17*	.22**	.40**	1.00			3.31	1.16
29. Freeze5	-.07	-.03	-.01	-.05	-.06	.11	-.01	-.07	.22**	.03	.23**	.15*	-.11	-.07	-.01	-.08	-.05	-.14	.21**	.16*	-.02	.02	.06	-.02	.10	.10	.18*	.21**	1.00		3.34	1.29
30. Freeze6	-.11	-.09	-.08	-.25**	-.10	.00	.02	.00	.13	.04	.20**	.22**	-.09	-.11	-.12	-.14	-.00	-.09	.14	.08	.00	.09	.15*	.14	.16*	.16*	.27**	.32**	.37**	1.00	2.53	1.18

Notes: All analysis provided were performed on raw data.
* $p < .05$
**$p < .01$
BAS = Behavioral activation system, BIS = Behavioral inhibition system

5.1.2 BAS – Behavioral Activation System

All 6 related BAS items loaded onto a latent factor when collapsed across sex (Table 1; factor loadings > 0.3 and $p < 0.01$). When estimating the parameter values across sex using a multigroup CFA, there was evidence of configural invariance as all the factor loadings were in the same direction and statistically significant (Table 1) where the standardized loadings ranged from 0.36-0.83, and the model fit the data well (see Table 3 for the configural model). When testing for metric invariance (equivalence in factor loadings), forcing the factor loadings to be equal across sex did not substantially influence model fit (see Table 3 for the metric model). The test for scalar invariance was rejected (equivalence in item intercepts); forcing item intercepts across sex to be equal resulted in a significant decrease in model fit (see Table 3 for scalar model). We then tested for partial scalar invariance by using nested models to probe specific items to identify the source of misfit. Items 3 and 5 were identified as having statistically distinct factor loadings across groups and by allowing the factor loadings to be different across groups, this resulted in a model that was able to account for the misfit and partial scalar invariance could be established (Table 3 for partial scalar model).

5.1.3 BIS – Behavioral Inhibition System

When considering BIS, all items loaded onto a factor when collapsing across sex but the multigroup CFA did not generate the same pattern of factor loadings across sex as all the factor loadings were not substantial and not significant (Table 1; standardized loadings were below 0.3 and p-values below 0.05). Items 1 and 6 were statistically significant in the female group but not the male group. Item 5 was also not statistically significant in either of the groups. As configural invariance could not be confirmed, we are unable to conduct any further analysis for measurement invariance and metric and scalar invariance could not be tested.

5.1.4 Fight

As displayed in Table 1, all Fight items loaded onto a factor when collapsing across sex and when estimating parameter values across sex using a multigroup CFA, we were able to confirm that the 6 items related to Fight loaded to a latent factor as all the factor loadings were substantial and significant (standardized loadings ranged from 0.56-0.79) and the model fit the data well (Table 4, configural model).

The test for metric invariance was rejected for sex (Table 4, metric model) as forcing factor loadings to be equal across sex resulted in a substantial decrement in model fit. Tests for partial metric invariance were conducted to investigate the source of misfit, however, this was unable to identify the source of misfit for metric Invariance in sex, and no partial metric invariance model was able to fit the data well. As scalar invariance testing requires full or partial metric invariance, rejecting the model for partial metric invariance precluded any further analysis for measurement invariance.

5.1.5 Flight

All Flight items loaded onto a factor when collapsing across sex but the multigroup CFA was not able to produce the same configuration of factor loadings across sex (Table 1; standardized loadings were below 0.3 and p-values below 0.05). Specifically, the factor loading for item 1 was not statistically significant in the female group and item 6 was not statistically significant in either group. As configural invariance must be confirmed prior to conducting analysis for metric or scalar invariance, rejecting the configural invariance hypothesis precluded us from conducting any further measurement invariance analysis.

Table 3 – Summary of tests of measurement invariance for BAS

	Model	χ^2 *(df)*	CFI	RMSEA (90% CI)	SRMR	Model Comparison	$\Delta\ \chi^2$ *(df)*	ΔCFI	ΔRMSEA	ΔSRMR	Decision
M1	Configural Invariance	22.37(16)	0.98	0.07 (0.0-0.12)	0.05	---	---	---	---	---	Accept
M2	Metric Invariance	30.49(21)	0.97	0.07 (0.0-0.12)	0.10	M1	8.17 (5)	0.01	0.004	0.053	Accept
M3a	Scalar Invariance	42.29(26)	0.95	0.08 (0.03-0.13)	0.11	M2	11.80 (5)	0.02	0.013	0.011	Reject
M3b	Partial Scalar Invariance	34.73(24)	0.97	0.07 (0.0-0.12)	0.11	M2	4.24 (3)	0.001	0.0	0.01	Accept

Note: Model 1 (M1) tests equivalence of model form (Configural Invariance); Model 2 (M2) tests equivalence of factor loadings (Metric Invariance); Model 3a (M3a) tests equivalence of item intercepts (Scalar Invariance); Model 3b has unconstrained item intercepts for BAS3 and BAS5.

Table 4 – Summary of tests of measurement invariance for Fight

	Model	χ^2 *(df)*	CFI	RMSEA (90% CI)	SRMR	Model Comparison	$\Delta\ \chi^2$ *(df)*	ΔCFI	ΔRMSEA	ΔSRMR	Decision
M1	Configural Invariance	24.11(18)	0.98	0.06 (0.0-0.12)	0.04	---	---	---	---	---	Accept
M2a	Metric Invariance	34.56(23)	0.96	0.07 (0.0-0.12)	0.09	M1	10.45 (5)	0.02	0.01	0.05	Reject
M2b	Partial Metric Invariance	31.42(19)	0.96	0.09 (0.02-0.14)	0.08	M1	7.31 (1)	0.02	0.03	0.04	Reject
M3	Scalar Invariance										

Note: Model 1 (M1) tests equivalence of model form (Configural Invariance); Model 2 (M2a) tests equivalence of factor loadings (Metric Invariance); Model 2b (M2b) has an unconstrained factor loadings for Fight 2, Fight 3, Fight 4, and Fight 5.

5.1.6 Freeze

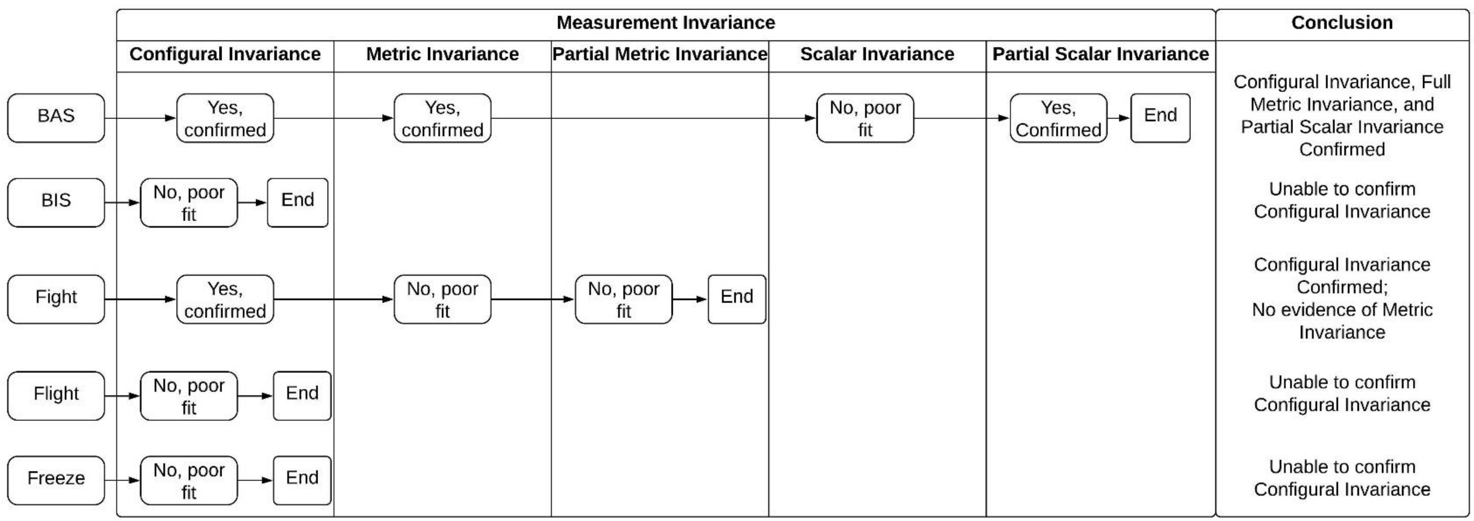
When we considered Freeze, all Freeze items did load onto a factor when collapsing across sex. However, the multigroup CFA did not produce the same configural of factor loadings across sex as all the factor loadings were not substantial and not significant (Table 1; factor loadings > 0.3 and $p < 0.01$). In particular, the factor loadings for items 1 and 5 were not statistically significant in the male group. As configural invariance is a prerequisite for metric and scalar invariance, we were unable to conduct any further measurement invariance analysis as the configural invariance hypothesis was rejected. A summary of the decision process for the Jackson-5 scales is provided (Figure 6).

5.1.7 Depression and Anxiety

We were able to confirm Anxiety as a latent factor using confirmatory factor analysis as all the loadings were substantial and significant and could test for measurement invariance. However, we were unable to confirm configural invariance for sex as the factor loading for DASS7 was below 0.03 in the group for sex; therefore, we were unable to test for measurement invariance.

We were able to confirm Depression did form a latent factor with confirmatory factor analysis as it met all the model fit requirements referenced. However, we were unable to test for measurement invariance. The items are treated as categorical and when testing for measurement invariance we are testing for equal responses across groups which we did not have for Depression as both groups have to have homogenous responses which we did not have for one or more of the items when testing for measurement invariance by sex.

Figure 6: Flow diagram for measurement invariance analytic decision process



5.2 Linear Regression

5.2.1 Fight

BAS maintained a positive relationship with Fight and remains constant in the model with the BAS-BIS interaction term, while BIS initially has a positive relationship with Fight. However, this becomes a negative relationship where the coefficient for BIS centers around zero in the model with an interaction term with BAS and BIS. The other coefficients included in the model (sex and age) remain consistent across all three models. In addition, the interaction term is not significant (Table 4).

5.2.2 Flight

BAS and BIS also have a positive relationship with Flight and this relationship does not change and even sees a stronger relationship with the model which includes a BAS-BIS interaction term. Notably, the BAS coefficient not only increases in the model with the BAS-BIS interaction term but becomes significant as well. The interaction term between BAS-BIS in the Flight model is significant which may indicate that an increase in BAS, which leads to an increase in Flight, also has a slightly lower total effect on the value for Flight when BIS is included in the model. The additional covariates added to the model for sex and age remain consistent in their coefficient value and relationship to Flight across all three models (Table 5).

Table 5 – Regression Coefficients for Fight

	Model 1: BAS Only			Model 2: BIS Only			Model 3: BAS-BIS Interaction		
	Coef	p-value	95% CI	Coef	p-value	95% CI	Coef	p-value	95% CI
(Intercept)	22.35	<0.01	(13.08, 31.63)	22.39	<0.01	(12.8,31.98)	23.05	<0.01	(11.22, 34.87)
BAS	0.25	<0.01	(0.11, 0.39)	-----------	-----------	-----------	0.07	0.77	(-0.43, 0.57)
BIS	-----------	-----------	-----------	0.2	0.02	(0.03, 0.36)	-0.02	0.94	(-0.54, 0.5)
BAS:BIS	-----------	-----------	-----------	-----------	-----------	-----------	0.01	0.55	(-0.02, 0.04)
Sex	-4.25	<0.01	(-5.59, -2.91)	-4.01	<0.01	(-5.38, 2.64)	-4.11	<0.01	(-5.46, -2.77)
Age	-0.24	0.32	(-0.7, 0.23)	-0.22	0.37	(-0.69, 0.26)	-0.25	0.3	(-0.71, 0.22)

Table 6 – Regression Coefficients for Flight

	Model 1: BAS Only			Model 2: BIS Only			Model 3: BAS-BIS Interaction		
	Coef	p-value	95% CI	Coef	p-value	95% CI	Coef	p-value	95% CI
(Intercept)	15.11	<0.01	(6.8, 23.43)	13.06	<0.01	(4.78, 21.34)	5.43	0.3	(-4.87, 15.72)
BAS	0.1	0.14	(-0.03, 0.22)	-----------	-----------	-----------	0.54	0.02	(0.1, 0.97)
BIS	-----------	-----------	-----------	0.22	<0.01	(0.08, 0.36)	0.71	<0.01	(0.26, 1.16)
BAS:BIS	-----------	-----------	-----------	-----------	-----------	-----------	-0.03	0.02	(-0.06, -0.01)
Sex	2.32	<0.01	(1.12, 3.52)	2.53	<0.01	(1.35, 3.71)	2.43	<0.01	(1.25, 3.6)
Age	-0.28	0.19	(-0.7, 0.14)	-0.29	0.17	(-0.7, 0.12)	-0.29	0.16	(-0.7, 0.12)

5.2.3 Freeze

BAS has a near zero relationship with Freeze as the confidence interval for the model with and without the BAS-BIS interaction centers around zero, while BIS has a positive relationship with Freeze in the model with and without the BAS-BIS interaction term. In the model with the BAS-BIS interaction term, the coefficient for BAS is also nonsignificant and the coefficient for BIS is significant. The BAS-BIS interaction term is also not significant and the additional covariates (sex and age) remain consistent with their relationship to Freeze between all three models (Table 6).

5.2.4 FFFS – Fight-Flight-Freeze-System

As illustrated in Table 7, BAS and BIS both have a positive and significant relationship with the FFFS and the effects are increased in magnitude in the model with the BAS-BIS interaction. However, in the BAS-BIS interaction model BAS is no longer significant, and the interaction term is not significant. The covariates, age and sex, do remain consistent in magnitude throughout all three models for the FFFS.

Table 7 – Regression Coefficients for Freeze

	Model 1: BAS Only			Model 2: BIS Only			Model 3: BAS-BIS Interaction		
	Coef	p-value	95% CI	Coef	p-value	95% CI	Coef	p-value	95% CI
(Intercept)	18.76	<0.01	(10.67, 26.85)	14.74	<0.01	(6.83, 22.65)	12.46	0.01	(2.56, 22.36)
BAS	-0.03	0.68	(-0.15, 0.1)	-----------	-----------	-----------	0.1	0.64	(-0.32, 0.52)
BIS	-----------	-----------	-----------	0.26	<0.01	(0.13, 0.4)	0.5	0.02	(0.06, 0.93)
BAS:BIS	-----------	-----------	-----------	-----------	-----------	-----------	-0.01	0.33	(-0.04, 0.01)
Sex	0.84	0.16	(-0.33, 2.0)	1.04	0.07	(-0.09, 2.17)	1.07	0.06	(-0.06, 0.93)
Age	-0.25	0.23	(-0.66, 0.16)	-0.28	0.16	(-0.67, 0.11)	-0.27	0.18	(-0.66, 0.12)

Table 8 – Regression Coefficients for FFFS

	Model 1: BAS Only			Model 2: BIS Only			Model 3: BAS-BIS Interaction		
	Coef	p-value	95% CI	Coef	p-value	95% CI	Coef	p-value	95% CI
(Intercept)	9.88	<0.01	(5.39, 14.37)	8.33	<0.01	(3.96, 12.7)	6.33	0.02	(0.87, 11.8)
BAS	0.1	<0.01	(0.03, 0.17)	-----------	-----------	-----------	0.17	0.16	(-0.06, 0.4)
BIS	-----------	-----------	-----------	0.19	<0.01	(0.12, 0.27)	0.28	0.02	(0.04, 0.52)
BAS:BIS	-----------	-----------	-----------	-----------	-----------	-----------	-0.01	0.34	(-0.02, 0.01)
Sex	-0.05	0.89	(-0.69, 0.6)	0.14	0.66	(-0.48, 0.76)	0.09	0.78	(-0.53, 0.71)
Age	-0.2	0.08	(-0.43, 0.02)	-0.21	0.06	(-0.42, 0.01)	-0.21	0.05	(-0.43, 0.0)

6.0 Discussion

6.1 Structural Equation Modeling

There has historically been a lack of consideration of sex differences in research and now the National Institute of Health has even required the examination of sex differences in funded research even if studying sex differences is not the main objective of the proposed research[49]. Therefore, it is increasingly becoming more important for researchers to consider analytical techniques to measure sex differences in their studies.

Some researchers may presume that the items in the 'Jackson-5' form a unidimensional and reliable latent factor and believe that the measure can be a useful tool to investigate sex differences. However, when these assumptions are tested, this study was unable to find evidence of measurement invariance across sex. In other words, we were unable to reliably assume that the items of the 'Jackson-5' r-RST could be evaluated equally across sex in the sample population for this study.

When we tested the latent structure of each of the five behavioral factors in the 'Jackson-5': BAS, BIS, Fight, Flight, and Freeze; only BAS and Fight demonstrated evidence that the scale is measuring the same factor across sex (configural invariance). There was evidence that the BAS scale had the same amount of "noise" or measurement error (metric invariance) and did have partial equivalence of mean differences in the latent construct (partial scalar invariance). However, the Fight scale did not show evidence that the amount of measurement error was the same across sex (metric invariance). These results may indicate that the BAS and Fight scales are not able to make valid inferences across sex in this study sample. Our analysis may also suggest that the latent structure of BIS, Flight, and Freeze respectively vary across sex and may not be measuring the same construct. Therefore, the 'Jackson-5' r-RST may not be a useful tool for investigating sex differences.

6.2 Linear Regression

The goal of replicating the SEM structure of the 'Jackson-5' using alternative methods became two-fold. The first is to attempt to verify the SEM factor structure as illustrated in Jackson's paper (Figure 1 in Jackson 2009)[37]. The second is to test if the alternative methods replicate the either our initial results using structural equation modeling where BAS is the only latent factor that could be confirmed through confirmatory factor analysis and showed evidence of scalar invariance across sex.

The first step is to reproduce the Fight-Flight-Freeze-System (FFFS) latent factor. In order to do this, we first created a series of linear regressions where each of the factors in the FFFS were the outcome variable and demographic variables (Sex, Age) along with BAS and BIS (separately and combined with interactions). If there was little change between the coefficients (direction, magnitude) then it could be argued that combining each of these scales to create the FFFS would be appropriate.

The results suggest that we did not have stable estimates across the regression models where Fight, Flight, and Freeze were the outcome variable. While the coefficients for Sex and Age remained consistent in magnitude and direction in the Fight, Flight, and Freeze models, BAS and BIS did not remain constant once they were covariates in the same model. In Fight and Freeze, the interaction term for BAS and BIS was not significant, meaning that the effects of BAS and BIS are unchanged in the same model. However, the interaction term for BAS and BIS was significant meaning that these covariates have a slightly lower total effect when added to the same model. Interestingly, BAS and BIS have an overall stronger effect in the Flight regression model when BAS and BIS are added in the same model. Normally, we would see a loss of power in a more complex model.

In the model were Fight, Flight, and Freeze are combined to form the FFFS outcome variable, we also do not have stable estimates across the various regression models. While Age, Sex, and BIS maintained a consistent regression coefficient and confidence interval, BAS did not. The interaction term was also not significant, so the effects of BAS and BIS are unchanged in the model where BAS and BIS are covariates in the same regression model. These results are consistent with the regression models where Fight, Flight, and Freeze were modeled separately as BAS and BIS did not have stable estimates across each of the components of the FFFS and in the combined FFFS model. These differences in coefficients may indicate that the structure of the FFFS is either unstable or the method of combining the items for each of the factors of the FFFS is not appropriate.

Notably, the FFFS has been criticized by others such as Krupic et al., which concluded the Fight scale is not correlated to the other factors in the FFFS which makes forming the FFFS inappropriate and has also claimed that Jackson's validation of this scale "show some theoretically ambiguous results"[50]. Therefore, the fact that we are seeing counterintuitive results may not be an artifact of the statistical method we selected but may highlight a problem with the development of the 'Jackson-5' scale. However, further analysis and multiple statistical methods should be considered prior to arriving at this conclusion.

The regression analysis relates to the methodology of structural equation modeling as many of the same components in the latter statistical technique can be represented in a regression model, provided we accept certain assumptions. In regression, we are unable to test the latent structure for each of the 'Jackson-5' behavioral factors, and we must assume that the summation across the quantiles of each questionnaire item reflects the true "score" for the BAS, BIS, Flight, Flight, Freeze, and FFFS respectively. However, this is not too dissimilar to how confirmatory factor analysis standardizes the items for each latent factor and utilizes matrix algebra to form a latent construct from each of the items.

By using this assumption, we were able to test the second order latent factor of the ‘Jackson-5’, the FFFS. This was unable to be performed in this study using confirmatory factor analysis as there are heavier restrictions or assumptions placed in this analytical method that are not applied in regression. However, investigators that prefer to use structural equation modeling in social and behavioral sciences research may not want to sacrifice these assumptions in favor of simpler regression modeling. Ultimately, it should still be left up to the investigators and collaborators to decide the appropriate analytical plan for their research.

6.3 Limitations

There are several limitations in the present study that should be acknowledged: 1) the study was conducted by a sample of undergraduate students that may not be generalizable to other populations, 2) the study oversampled students with a history of a major depressive episode, and 3) the conclusions of this study are only specific to the current sample and should not be generalized.

7.0 Conclusions/Recommendations

In this study, we have determined that there are sex differences in the psychometric properties of the 'Jackson-5' scale which may stress the importance of ensuring other behavioral or psychometric scales are measuring the same construct across sex. The results of this study may suggest that the 'Jackson-5' scale may not be a useful tool for investigating sex differences. This highlights the need to consider measuring sex differences or across groups by stringent statistical analytical techniques – such as structural equation modeling or item response theory – as these techniques allow for testing the quality of measurement across groups and the mean level differences across groups. We were also unable to confirm the latent structure of the 'Jackson-5' through regression analysis, however additional analytical techniques or strategies should be considered before arriving at this conclusion.

References

1. Felton JW, Collado A, Shadur JM, Lejuez CW, MacPherson L. Sex differences in self-report and behavioral measures of disinhibition predicting marijuana use across adolescence. Exp Clin Psychopharmacol 2015;23:265-74.
2. Kampmann IL, Emmelkamp PMG, Morina N. Self-report questionnaires, behavioral assessment tasks, and an implicit behavior measure: do they predict social anxiety in everyday life? PeerJ 2018;6:e5441.
3. Wiegand DL, Wilson T, Pannullo D, et al. Measuring Acute Pain Over Time in the Critically Ill Using the Multidimensional Objective Pain Assessment Tool (MOPAT). Pain Manag Nurs 2018;19:277-87.
4. Wong QJ, Moulds ML. The relationship between the maladaptive self-beliefs characteristic of social anxiety and avoidance. J Behav Ther Exp Psychiatry 2011;42:171-8.
5. Wong QJ, Moulds ML. Impact of rumination versus distraction on anxiety and maladaptive self-beliefs in socially anxious individuals. Behav Res Ther 2009;47:861-7.
6. Vergara-Lopez C, Lopez-Vergara HI, Colder CR. Executive functioning moderates the relationship between motivation and adolescent depressive symptoms. Pers Individ Dif 2013;54:18-22.
7. Lopez-Vergara HI, Colder CR, Hawk LW, Jr., et al. Reinforcement sensitivity theory and alcohol outcome expectancies in early adolescence. Am J Drug Alcohol Abuse 2012;38:130-4.
8. Hennegan JM, Loxton NJ, Mattar A. Great expectations. Eating expectancies as mediators of reinforcement sensitivity and eating. Appetite 2013;71:81-8.
9. Brown TA. Confirmatory Factor Analysis for Applied Research: The Guilford Press; 2015.
10. Bollen K. Structural equations with latent variables.: John Wiley & Sons; 1989.
11. Jöreskog KG, Sörbom D. LISREL 10 for Windows. Skokie, IL: Scientific Software International, Inc; 2018.
12. SAS Institute. The SAS system for Windows. 9.4 ed2019.
13. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.
14. Muthén LK, Muthén BO. Mplus User's Guide. Los Angeles, CA: Muthén & Muthén; 1998-2017.
15. van de Schoot R, Lugtig P, Hox J. Developmetrics A checklist for testing measurement invariance. European Journal of Developmental Psychology 2012;9.
16. Putnick DL, Bornstein MH. Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. Dev Rev 2016;41:71-90.
17. Mair P. Modern Psychometrics with R. 1 ed: Springer International Publishing; 2018.
18. Little TD. Longitudinal Structural Equation Modeling: Guilford Press; 2013.
19. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. J Appl Psychol 2008;93:568-92.
20. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research. International Journal of psychological research 2010;3:111-30.
21. Muthen B, Asparouhov T. Recent Methods for the Study of Measurement Invariance with Many Groups: Alignment and Random Effects. Sociological Methods & Research 2018;47:637-64.

22. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research International Journal of Psychological Research 2010;3:111-21.
23. Dimitrov DM. Testing for Factorial Invariance in the Context of Construct Validation. Measurement and Evaluation in Counseling and Development 2010;43:121-49.
24. Steinmetz H. Analyzing observed composite differences across groups. Methodology 2013.
25. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental review 2016;41:71-90.
26. Hoffman L. Multiple Group CFA Invariance Example (data from Brown Chapter 7) using MLR Mplus 7.4: Major Depression Criteria across Men and Women (n = 345 each)2018.
27. Gray JA. The neuropsychology of anxiety : an enquiry into the functions of the septo-hippocampal system. Oxford New York: Clarendon Press; Oxford University Press; 1982.

28. Gray JA, McNaughton N. The neuropsychology of anxiety : an enquiry into the functions of the septo-hippocampal system. 2nd ed. Oxford ; New York: Oxford University Press; 2000.
29. Corr PJ. J. A. Gray's reinforcement sensitivity theory and frustrative nonreward: a theoretical note on expectancies in reactions to rewarding stimuli. Personality and Individual Differences 2002;32:1247-53.
30. Vergara-Lopez C, Lopez-Vergara HI, Roberts JE. Testing a "content meets process" model of depression vulnerability and rumination: Exploring the moderating role of set-shifting deficits. J Behav Ther Exp Psychiatry 2016;50:201-8.
31. Hundt NE, Kimbrel NA, Mitchell JT, Nelson-Gray RO. High BAS, but not low BIS, predicts externalizing symptoms in adults. Pers Individ Dif 2008;44:565-75.
32. Donahue JJ, Caraballo LJ. Examining the triarchic model of psychopathy using revised Reinforcement Sensitivity Theory. Personality and Individual Differences 2015;80:125-30.
33. Nob RM. Examining the triarchic model of psychopathy using revised Reinforcement Sensitivity Theory. Educational Measurement and Evaluation Review 2013;4:26-36.
34. Smillie LD, Pickering AD, Jackson CJ. The new reinforcement sensitivity theory: implications for personality measurement. Pers Soc Psychol Rev 2006;10:320-35.
35. Kramer SL, Rodriguez BF. A Comparison of Reinforcement Sensitivity Theory Measures: Unique Associations With Social Interaction Anxiety and Social Observation Anxiety. 2018 2018;25:627-39.
36. Smillie LD, Jackson CJ. Functional impulsivity and reinforcement sensitivity theory. J Pers 2006;74:47-84.
37. Jackson CJ. Jackson-5 scales of revised Reinforcement Sensitivity Theory (r-RST) and their application to dysfunctional real world outcomes. Journal of Research in Personality 2009;43:556-69.
38. Walker BR, Jackson CJ. Examining the validity of the revised Reinforcement Sensitivity Theory scales. Personality and Individual Differences 2017;106:90-4.
39. Pagliaccio D, Luking KR, Anokhin AP, et al. Revising the BIS/BAS Scale to study development: Measurement invariance and normative effects of age and sex from childhood through adulthood. Psychol Assess 2016;28:429-42.

40. Vervoort L, De Caluwe E, Vandeweghe L, et al. Parent-Reported BIS/BAS Scales for Children: Factor Structure and Measurement Invariance Across Age and Gender. Assessment 2019;26:1282-95.
41. Afsahri B, Rasouli-Azad M, Ghoreishi FS. Comparison of original and revised reinforcement sensitivity theory in clinically-stable schizophrenia and bipolar disorder patients. Personality and Individual Differences 2018:321-7.
42. Beuckelaer A, Swinnen G. Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. In: Davidov E, Schmidt P, Billiet J, Meuleman B, eds. Cross-cultural analysis : methods and applications. 2nd edition. ed. New York: Taylor and Francis; 2018.
43. First MB, Spitzer R, Gibbon M, Williams J. Structured Clinical Interview for DSM-IV axis I Disorders, Clinical Version (SCID-CV). Washington, D. C.: American Psychiatric Press, Inc; 1996.
44. Ferrer M, Andion O, Calvo N, et al. Clinical Components of Borderline Personality Disorder and Personality Functioning. Psychopathology 2018;51:57-64.
45. Peirce JM, Burke CK, Stoller KB, Neufeld KJ, Brooner RK. Assessing traumatic event exposure: comparing the Traumatic Life Events Questionnaire to the Structured Clinical Interview for DSM-IV. Psychol Assess 2009;21:210-8.
46. Altunoz U, Kokurcan A, Kirici S, Bastug G, Ozel-Kizil ET. Clinical characteristics of generalized anxiety disorder: older vs. young adults. Nord J Psychiatry 2018;72:97-102.
47. Farb N, Anderson A, Ravindran A, et al. Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. J Consult Clin Psychol 2018;86:200-4.
48. Hallquist MN, Wiley JF. MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. Struct Equ Modeling 2018;25:621-38.
49. Clayton JA. Applying the new SABV (sex as a biological variable) policy to research and clinical care. Physiol Behav 2018;187:2-5.
50. Krupic D, Corr PJ, Rucevic S, Krizanic V, Gracanin A. Five reinforcement sensitivity theory (RST) of personality questionnaires: Comparison, validity and generalization. Personality and Individual Differences 2017;97:19-24.

Appendix A: Mplus Tutorial for Confirmatory Factor Analysis and Measurement Invariance

Mplus - [CFA_BAS_example]

File Edit View Mplus Plot Diagram Window Help

```
TITLE:

DATA:

FILE IS 07112018.xls;


VARIABLE:

NAMES ARE
Gender Age Hx
BAS1 BAS2 BAS3 BAS4 BAS5 BAS6
BIS1 BIS2 BIS3 BIS4 BIS5 BIS6
Fight1 Fight2 Fight3 Fight4 Fight5 Fight6
Flight1 Flight2 Flight3 Flight4 Flight5 Flight6
Freeze1 Freeze2 Freeze3 Freeze4 Freeze5 Freeze6;

USEVARIABLES ARE
BAS1 BAS2 BAS3 BAS4 BAS5 BAS6;

MISSING are all (-9999999);

ANALYSIS:

MODEL:

BAS_f BY BAS1* BAS2 BAS3 BAS4 BAS5 BAS6;
BAS_f@1;

BAS2 WITH BAS1;
```

The CFA model is specified as follows:

Latent Factor BY Item1* Items2-N

The first item is followed by an (*) to standardize the other loadings to the same scale.

Mod indices (as previously described) can also be added between items using a WITH command:

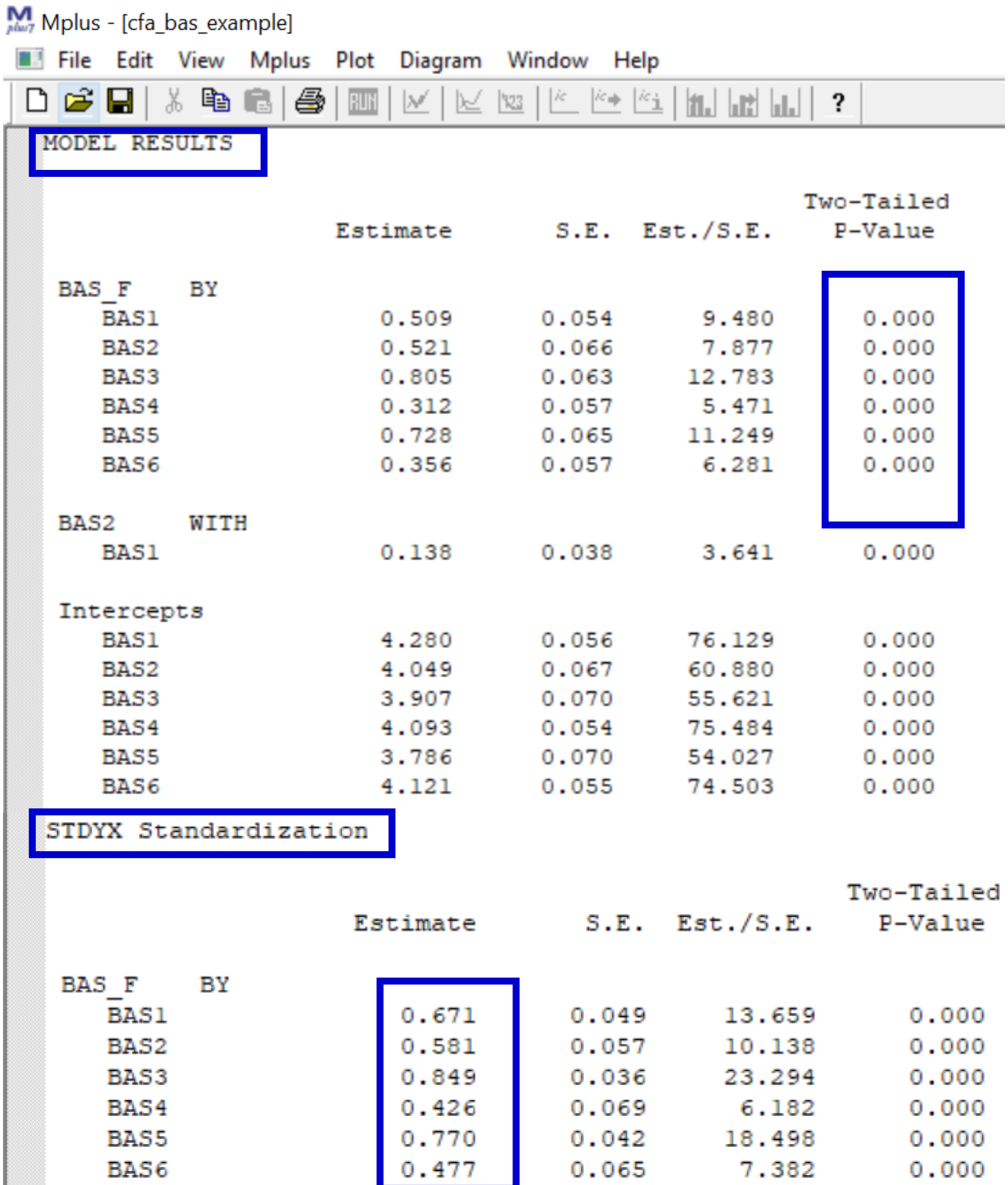
ItemX1 WITH ItemX2

Mplus - [cfa_bas_example]

File Edit View Mplus Plot Diagram Window Help

```
MODEL FIT INFORMATION

Number of Free Parameters                        19

Loglikelihood

          H0 Value                        -1177.794
          H1 Value                        -1167.198

Information Criteria

          Akaike (AIC)                     2393.587
          Bayesian (BIC)                   2454.463
          Sample-Size Adjusted BIC         2394.288
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              21.192
          Degrees of Freedom                      8
          P-Value                            0.0067

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.095
          90 Percent C.I.                     0.047  0.146
          Probability RMSEA <= .05            0.060

CFI/TLI

          CFI                                 0.962
          TLI                                 0.929
```

All Mplus run SEM models will include model fit information such as:, CFI, Chi-Square, RMSEA, and others



Mplus - [cfa_bas_example]

File Edit View Mplus Plot Diagram Window Help

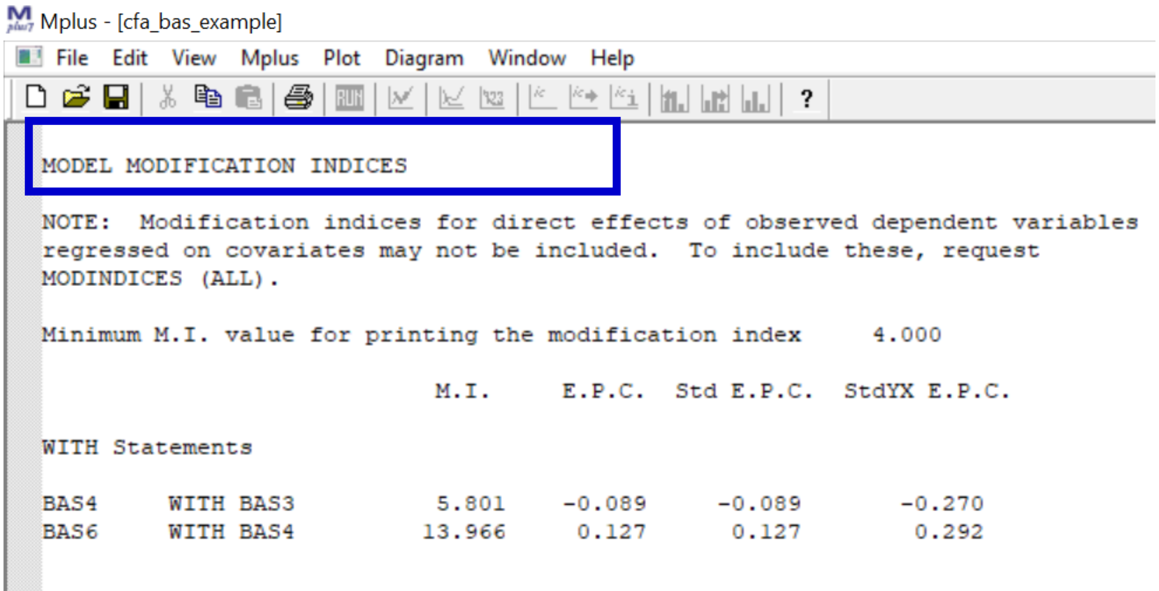
MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
BAS_F BY				
BAS1	0.509	0.054	9.480	0.000
BAS2	0.521	0.066	7.877	0.000
BAS3	0.805	0.063	12.783	0.000
BAS4	0.312	0.057	5.471	0.000
BAS5	0.728	0.065	11.249	0.000
BAS6	0.356	0.057	6.281	0.000
BAS2 WITH				
BAS1	0.138	0.038	3.641	0.000
Intercepts				
BAS1	4.280	0.056	76.129	0.000
BAS2	4.049	0.067	60.880	0.000
BAS3	3.907	0.070	55.621	0.000
BAS4	4.093	0.054	75.484	0.000
BAS5	3.786	0.070	54.027	0.000
BAS6	4.121	0.055	74.503	0.000

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
BAS_F BY				
BAS1	0.671	0.049	13.659	0.000
BAS2	0.581	0.057	10.138	0.000
BAS3	0.849	0.036	23.294	0.000
BAS4	0.426	0.069	6.182	0.000
BAS5	0.770	0.042	18.498	0.000
BAS6	0.477	0.065	7.382	0.000

The model that was specified presents two versions of the model estimates: nonstandardized and standardized. Here, we only need to look at the p-value for the nonstandardized model. In the standardized model, we can verify the factor loadings (Estimate in Mplus).




```
MODEL MODIFICATION INDICES

NOTE:  Modification indices for direct effects of observed dependent variables
regressed on covariates may not be included.  To include these, request
MODINDICES (ALL).

Minimum M.I. value for printing the modification index     4.000

                              M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

WITH Statements

BAS4     WITH BAS3           5.801    -0.089    -0.089      -0.270
BAS6     WITH BAS4          13.966     0.127     0.127       0.292
```

As previously described, mod indices can be implemented to improve model fit. Mplus has an automatic output that requires no additional code or analysis. However, it is important not to overfit the model by adding mod indices unnecessarily.

Mplus - [BAS_measurement_invariance_configural invariance_gender]

File Edit View Mplus Plot Diagram Window Help

```
ANALYSIS:

!first model command corresponds to first group;
MODEL:

!factor loadings freely estimated, just labeled
BAS_F BY BAS1* (L1)
         BAS2* (L2)
         BAS3* (L3)
         BAS4* (L4)
         BAS5* (L5)
         BAS6* (L6);


!item intercepts all freely estimated, just label
[BAS1*] (i1);
[BAS2*] (i2);
[BAS3*] (i3);
[BAS4*] (i4);
[BAS5*] (i5);
[BAS6*] (i6);

!residual covariance
 BAS2 WITH BAS1* (C1);

!factor variance fixed to 1 for identification
BAS_f@1;

!factor mean fixed to 0 for identification
[BAS_f@0];
```

```
MODEL FEMALE:

!factor loadings f
BAS_F BY BAS1*
         BAS2*
         BAS3*
         BAS4*
         BAS5*
         BAS6*;

!item intercepts a
[BAS1*];
[BAS2*];
[BAS3*];
[BAS4*];
[BAS5*];
[BAS6*];

 !residual covaria
 BAS2 WITH BAS1*;

!factor variance f
BAS_f@1;

!factor mean fixed
[BAS_f@0];
```

The Configural Invariance model is specified same as the CFA model.

Here, we are assessing the latent factor across sex, so we need to specify the model twice, once for each group. The first model corresponds to males, and the second model corresponds to females.

We are not constraining the factor loadings or intercepts, so we are only going to look to see if the two groups still form a latent factor individually.

We are also setting the factor variance to 1 and the factor mean to 0.

Mplus - [BAS_measurement_invariance_metric invariance_gender]

File Edit View Mplus Plot Diagram Window Help

```
[BAS_f@0];

MODEL FEMALE:

!factor loadings NOW constrained to group 1
BAS_F BY BAS1* (L1)
         BAS2* (L2)
         BAS3* (L3)
         BAS4* (L4)
         BAS5* (L5)
         BAS6* (L6);
!item intercepts all freely estimated, not labeled
[BAS1*];
[BAS2*];
[BAS3*];
[BAS4*];
[BAS5*];
[BAS6*];

  !residual covariance
   BAS2 WITH BAS1*;

!factor variance can now be freely estimated
BAS_f*;

!factor mean fixed to 0 for identification
[BAS_f@0];
```

The Metric Invariance model is specified same as the Configural Invariance model with a few key differences.

Here, we are assessing whether the factor loadings are measured the same across sex. The first model specified for Males does not change. In the second model for Females, we now constrain the factor loadings to be the same as Males. We can do this by adding the same labels for factor loadings we used for Males for Females.

We are also setting the factor variance to be freely estimated but still set the factor mean to 0.

Mplus - [BAS_measurement_invariance_scalar invariance_gender]

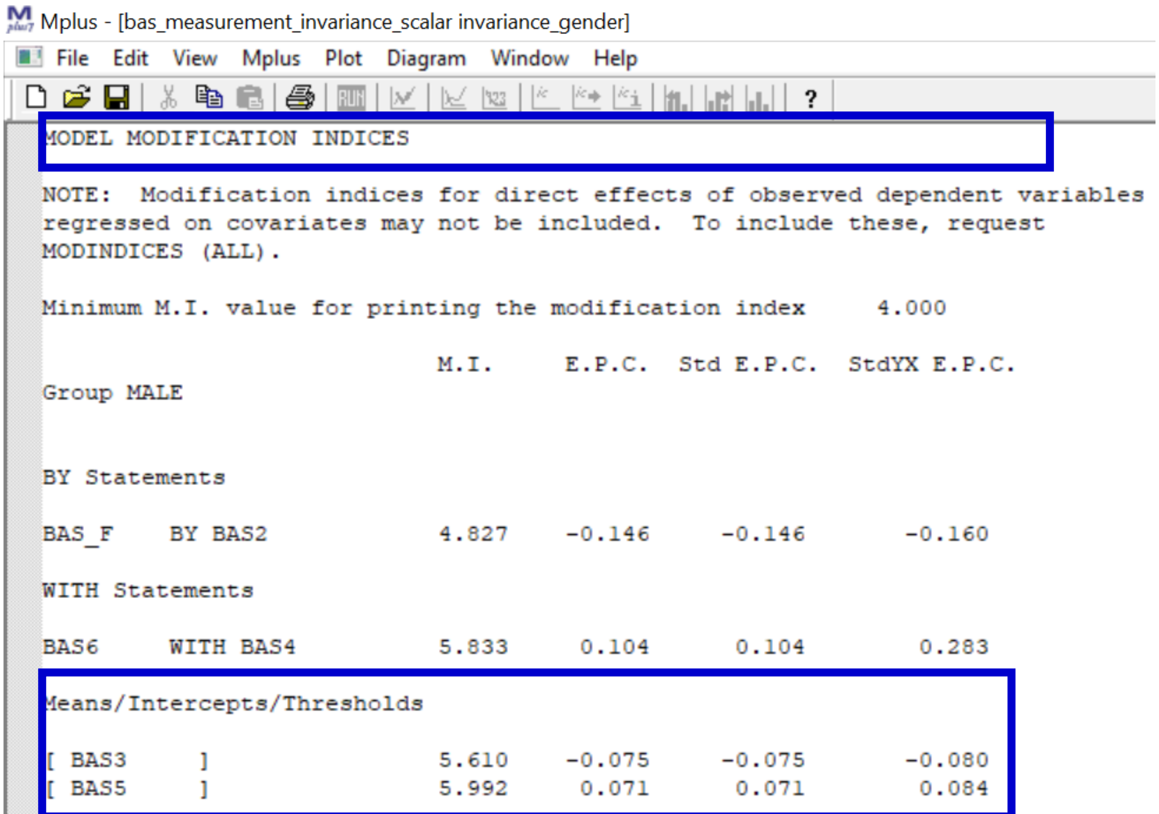
File Edit View Mplus Plot Diagram Window Help

```
[BAS_f@0];

MODEL FEMALE:

!factor loadings NOW constrained to group 1
BAS_F BY BAS1* (L1)
         BAS2* (L2)
         BAS3* (L3)
         BAS4* (L4)
         BAS5* (L5)
         BAS6* (L6);

!item intercepts NOW constrained to that of group 1
[BAS1*] (i1);
[BAS2*] (i2);
[BAS3*] (i3);
[BAS4*] (i4);
[BAS5*] (i5);
[BAS6*] (i6);

     !residual covariance
     BAS2 WITH BAS1*;

!factor variance is still freely estimated
BAS_f*;

!factor mean can now be freely estimated
[BAS_f*];
```

The Scalar Invariance model is specified same as the Metric Invariance model with a few key differences.

Here, we are assessing whether the item intercepts are measured the same across sex. The first model specified for Males does not change. In the second model for Females, we now constrain the factor loadings and the item intercepts to be the same as Males.

We are also setting the factor variance and the factor mean to be freely estimated.



If the fit of the model is substantially different going from either Configural to Metric Invariance or from Metric to Scalar Invariance, Mplus has a nice feature to identify the source of misfit and makes finding a partial invariance model very easy.

In this example, we need to identify the source of misfit going from Metric to Scalar invariance. That means that some of the item intercepts should not be constrained across groups or it will affect the model fit substantially.

We can look under the Modification Indices section and see that item intercepts BAS3 and BAS5 are the source of misfit. We can fit a partial Scalar Invariance model and all we need to change is remove the label for the item intercepts BAS3 and BAS5.