Reinforcement Sensitivity Theory of Personality Questionnaire: Measurement and Structural Invariance Across Age and Gender Groups

R. Gomez[a], S Watson[a,] J. Van Wynen[b], V. Stavropoulos[bc] and P. J. Corr[d].

[a]Federation University, Australia; [b]Cairnmillar Institute, Australia; [c]Victoria University, Australia; [d]City, University of London, UK.

Correspondence concerning this article should be addressed to Rapson Gomez. email: rapson.gomez@federation.edu.au

Abstract

The study examined the measurement (configural, metric, scalar, and residual) and structural (factor variances, covariances) invariance of the Reinforcement Sensitivity Theory of Personality Questionnaire (RST-PQ; Corr & Cooper, 2016) across gender and age groups for an ESEM version of the theorized six-factor oblique model. Multiple-group confirmatory factor analysis (CFA) supported full measurement and structural invariance. Scalar invariance was also supported by multiple indicators multiple causes (MIMIC) procedures that controlled for the effects of age and gender as appropriate. There was also no difference for the six latent mean scores across gender and age. The psychometric and practical implications of the findings are discussed.

Key Words: Reinforcement Sensitivity Theory of Personality Questionnaire, measurement invariance, males versus females

Reinforcement Sensitivity Theory of Personality Questionnaire: Measurement and Structural Invariance Across Age and Gender Groups

The revised version of reinforcement sensitivity theory (r-RST) is a neuropsychological model of personality (Corr & McNaughton, 2012). For measuring the constructs in this model, Corr and Cooper (2016) developed the Reinforcement Sensitivity Theory of Personality Questionnaire (RST-PQ), which has six scales, and it proposed structure is a six-factor oblique model. A recent study revealed more support for the exploratory structural equation modeling (ESEM) version than a confirmatory factor analysis (CFA) version of this model (Gomez et al., 2020), suggesting???not sur if needed, but…. The current study presents additional psychometric data for the RST-PQ model. More specifically, it provides new findings for measurement and structural invariance across males and females, and emerging adults and adults. 'emerging adults' – who are these people?

The RST-PQ has 65 self-rating items that are grouped into six scales (factors) corresponding to the r-RST components (Corr & Cooper, 2016). There are scales to measure the fight-flight-freeze system (FFFS), the behavioral inhibition system (BIS), and the behavioral approach system (BAS). The BAS scale includes subscales for Reward Interest (RI), Goal-Drive Persistence (GDP), Reward Reactivity (RR), and Impulsivity (I). As the RST-PQ is evolving to be a promising measure for studies involving r-RST (e.g., Bacon, Corr, & Satchell, 2018; Beaton, Mutinelli, & Corr, 2017; Jiang & Tiliopoulos, 2014), it is important that we have a good understanding of its psychometric properties.

Although a number of CFA studies on the factor structure of the RST-PQ have concluded support for the theorized six-factor oblique model (e.g., Corr and Cooper, 2016; Eriksson, Jansson, & Sundin, 2019; Krupić, Corr, Ručević, Križanić, & Gračanin, 2016; Pugnaghi, Cooper, Ettinger, & Corr, 2018; Wytykowska, Fajkowska, Domaradzka, & Jankowski, 2017), Gomez, Watson, Van Wynen, Trawley, Stavropoulos, and Corr (2020) have noted that the finding in past studies can, at best, be interpreted as showing mixed support, based on currently accepted validated fit index cutoffs proposed by Hu and Bentler (1999). According to Hu and Bentler cutoff levels for good model fit are root mean square error of approximation (RMSEA) $\leq .06$, and comparative fit index (CFI) $\geq .95$. Gomez et al. have pointed out that in virtually all past CFA studies of the RST-PQ, the RMSRA values showed good or acceptable fit, while the CFI showed poor fit. Gomez et al. conducted their own factor analysis of the RST-PQ. In addition to CFA, that study also used the more advanced exploratory structural equation modeling (ESEM) – an approach not used previously to study the factor structure of the RST-PQ. Results revealed good and most support for an ESEM version of the theorized six-factor model than the originally proposed six-factor CFA model. The support for the ESEM six-factor model is consistent with the six-factor structure of the RST-PQ. Although the proposed structure has been supported, it is argued here that we still lack other important psychometric information for an unbiased interpretation of the scores derived from this measure. Most notably there is limited and confusing data on measurement and structural invariance for the RST-PQ across gender and age groups. Before we discuss this matter, it would be prudent to explain the statistical concept of invariance, how it is examined, and the importance and relevance of testing invariance across gender and age groups for the RST-PQ.

Measurement invariance means that individuals in different groups who have the same latent score will endorse the same observed score on a measure (Reise, Widaman, & Paugh, 1993). This means that if there is weak no invariance the groups cannot be justifiably compared as the scores are confounded by different measurement and scaling properties. Expressed differently, invariance for a measure across groups being compared is a prerequisite for valid comparison of the groups. Thus, when males and females are compared on a questionnaire, the questionnaire has to have invariance across males and females for such a comparison to be valid. When applied to the RST-PQ, measurement invariance means that its scores need to have measurement invariance across groups (for example, males and females) in the first instance for the RST-PQ to be accurately used to differentiate these groups.

A widely used approach for testing measurement invariance is multiple-group CFA (Meredith, 1993; Vandenberg & Lance, 2000). Generally, with this approach four levels of measurement invariance are examined. There are configural, metric, scalar and residual invariance. Configural invariance tests if the same factor structure holds for the groups in question. Metric invariance tests if the factor loadings (strength of the associations between items and their latent factors) of like items are the same across the groups. Support for metric invariance means that it is appropriate to compare the groups for factor correlations and correlations of the factors with other constructs. Scalar invariance tests for equivalency in item intercepts across the groups. It indicates if members in the different groups with a certain latent score will endorse the same observed scores. It is a prerequisite for comparing the mean latent scores of the groups. Residual invariance tests for equivalency across the groups for measurement error, and it is also a prerequisite for comparing the mean observed scores across the groups. Metric, scalar and error variances invariances are often referred to as *weak*, *strong* and *strict* invariance (Meredith, 1993). When there is some support (full or partial) for measurement invariance, structural invariance (invariance for variances and covariances) can be tested and, also, the equivalency for latent means can be examined.

In a recent meta-analysis of studies that examined measurement invariance of personality measures across gender, Dong and Dumas (2020) concluded that while all studies supported configural invariance (29 out of 29) , 4 out of 29 studies (13.79%) did not support full metric invariance, and as many as 13 out of 29 studies (44.83%) did not support full scalar invariance). The same study also reported that for measurement invariance across age, all studies (26 out of 26) supported configural invariance. However, 4 out of 26 studies (15.38%) did not support full metric invariance, 41.18% did not support full scalar invariance, and most of studies did not support residual invariance. Although this meta-analysis focused on personality measures in general, it can be speculated that the findings highlight the need to demonstrate measurement invariance for personality measures, such as the RST-PQ, especially for groups often used in research studies, such as gender and age groups. In this respect, as both gender and age groups can influence responses to personality measures, it will be useful to control the effects of age when examining invariance across gender; and to control the effects of gender when examining invariance across age. The can be done using the multiple indicators multiple causes models (MIMIC; Joreskog & Goldberger, 1975).

MIMIC models can test if members belonging to different groups endorse the same levels of the item after being equated on the underlying latent trait measuring the item. This test is comparable to testing scalar invariance in the multiple-group CFA approach. When there is no equivalency (members in different groups endorsing different levels of the item when they have for the same level of the underlying latent trait), the item is said to have differential item functioning (DIF). A feature of the MIMIC approach that is important for the current study is that it can control the possible confounding effects of other variables (or covariates), such as age when examining DIF as a function of gender, or gender when examining DIF as a function of age.

To date, as far as we are aware, there have been at least two studies that have examined invariance across gender and age groups (Eriksson, Jansson, & Sundin, 2019; Pugnaghi, Cooper, Ettinger, & Corr, 2018). The study by Eriksson et al. used multiple-group CFA to examine invariance across young adults (< 44 years) and older adults (> 44 years); and females and males for a revised version of the RST-PQ (with 52 items rather than the original 65 items), with a six-factor CFA model as the structural model. The findings across the age groups indicated support for configural, and full metric and residual invariance. For scalar invariance, six items showed lack of invariance. They were BIS items 10, 17, 21, and 57, RR item 30, and GDP item 31. For gender groups, there was also support for configural, and full metric and residual invariance. For scalar invariance, only a single item (item 16) showed lack of invariance. The study by Pugnaghi et al used the MIMIC approach for the theorized six-factor CFA model. The authors found support for measurement invariance across gender groups and for age. Overall, therefore, although we have data on measurement invariance across gender and age for the RST-PQ, it is limited and inconsistent. Indeed, only one study has examined invariance for the complete version of the RST-PQ, and no study has examined structural invariance for the RST-PQ, or the measurement and structural invariance for RST-PQ in terms of the six-factor ESEM model that was proposed as the preferred RST-PQ model (Gomez et al., 2020).

Given existing limitations and omissions, the aim of the current study was to examine invariance across ratings of adult gender and age groups for the RST-PQ. The age groups were emerging adults (18 years to 29 years) and adults (≥ 30 years). We decided on these age groups as emerging adulthood is now recognized as a unique stage of development, different from older and younger age groups in many areas. Among other differences, there is increased instability in personality, such as in social dominance, conscientiousness, and emotional stability (Arnett, 2004). We also wish to point out at the outset that the RST-PQ ratings (data set) used in the current study was the same data set used in the previous factor analysis study of this measure (Gomez et al., 2020). As will be recalled, that study demonstrated most support for the six-factor ESEM model. Consequently we used this model as the factor model in all our analyses in the current study. We first used multiple-group ESEM with target rotation to test measurement (configural, metric, scalar, and residual) and structural (variances and covariances) invariance, across gender (male v female), and across age (emerging adults v adults) groups. Following this, we used an ESEM based MIMIC model to test for DIF across males and females, controlling for age; and across emerging adults and adults, controlling for gender. In both the multiple-group CFA and MIMIC analyses, we also examined differences in latent mean scores across the groups in question. Methodologically, the multiple-group ESEM and ESEM based MIMIC approaches are closely comparable to their CFA counterparts. The difference is that, unlike a CFA model where cross-loadings are not allowed (constrained to zero), in ESEM, items loaded on the designated factors, and cross-loadings are "targeted," but not forced, to be as close to zero as possible.

Method

Participants

All data were collected from participants residing in Australia. This sample (N =901) comprised 672 females (743.6%) and 229 males (25.4%). Age ranged from 18 to 82 years (M = 32.07, SD = 16.38). The mean (SD) age for females and males were 32.44 years (16.22 years) and 30.99 years (16.83 years), respectively. Females and males did not differ significantly on age, t (df = 899) = 1.15, p = 0.11. The number of participants classified as emerging adults and adults were 588 (females = 427 and males = 161) and 313 (females = 245 and males = 68), respectively. The groups did not differ for gender distribution, χ^2 (df = 1) = 3.45, p = .065. The majority of participants (56%) were students recruited from the psychology participant pool in exchange for course credit points. Other participants were members of the general community, recruited mainly through paid advertisements posted on social media (Facebook). Although details are not shown, the majority of participants were working, had completed secondary school education, and was in some sort of relationship with a partner.

Measure

All participants completed a demographic sheet that sought information about their age gender, education, employment and relationship status; and also the RST-PQ. The RST-PQ, a self-report questionnaire for measuring the constructs in r-RST, was described briefly in the introduction. It has subscales for: FFFS (10 items); BIS (23 items); and BAS-Reward Interest (RI; 7 items), BAS-Goal-Drive Persistence (GDP; 7 items), BAS-Reward Reactivity (RR; 10 items), and BAS-Impulsivity (I; 8 items). All items are rated on a four-point scale, ranging from 1 (*not at all*) to 4 (*highly*). Corr and Cooper (2016) reported Cronbach's alpha values of FFFS =.78, BIS = .93, RI =.75, GDP = 86, RR = .78, I = .74 in the initial development and validation study of the RST-PQ. The Cronbach's alpha values for the current study were FFFS =.79, BIS = .91, RI =.74, GDP = .87, RR = .75, I = .75.

Procedure

The data for the study were collected in the State of Victoria, Australia. Ethics approval for the recruitment of participants was obtained from the Human Research Ethics Committee of Federation University Australia, and the Cairnmillar Institute Human Research Ethics

Committee. All participants were recruited by advertisements and on-line (via Survey Monkey), and they were initially provided with an information statement providing sufficient about the study to enable that to make an informed decision about participation. Consenting participants completed the questionnaires anonymously, and were not compensated for their participation.

Statistical Analysis

In terms of statistical power, the sample size (N = 901) in the current study is well above the level recommended by some researchers for the factor analyses involving 65 indicator items (i.e., a minimum sample size of 65 x 10 = 650; Myers, et al., 2011). The ESEM model in the study were conducted using geomin (i.e., oblique) rotation. As mentioned earlier, in ESEM, items loaded on the designated factors, and cross-loadings were "targeted," but not forced, to be as close to zero as possible.

The steps for testing invariance in in the multiple-group ESEM approach is similar to that used in the multiple-group CFA approach, described by others (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). In brief, this procedure involves comparing progressively more constrained models that test for measurement invariance: configural invariance (equality for form), metric (weak) invariance (equality for factor loadings), scale (strong) invariance (equality for responses to items), and error variances (strict) invariance (equality for uniqueness). When there is some support for measurement invariance (full or partial), structural invariance (equivalency for factor variances and covariances) can be examined. Additionally, the groups can be compared for latent mean scores, taking into account the non-invariance in the measurement model. As the details involved in the different step have been extensively covered in the literature (e.g., Meredith, 1993; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000), details are not been provided here.

The steps for testing invariance in the ESEM-based MIMIC approach are similar to that used in the CFA-based MIMIC approach. DIF items were identified using a procedure similar to that used by Gomez and Vance (2008). In this procedure, initially, a baseline model with all paths from the predictor to all items constrained to zero (M1) is computed. Following this, the item with the highest modification index between the predictor and the items is identified. The path of this item is then freely estimated in a subsequent model (revised model). A difference in fit between the revised model and the initial model is indicative of DIF in the freed item (Muthén, 1988; see also Brown, 2006). This process is repeated until all DIF items are identified (Muthén, 1988; see also Brown, 2006). In the MIMIC model for testing DIF across the gender groups, the RST-PQ items loaded on their respective latent factors, and these factors were regressed on gender (the predictor), with correlations between gender and age (the covariate). Thus the effect of age was controlled in the analysis. In the MIMIC model for test DIF across the age groups, the RST-PQ items loaded on their respective latent factors, and these factors were regressed on age (the predictor), with correlations between age and gender(the covariate). Thus gender was controlled in the analysis.

All statistical analyses were conducted using M*plus* Version 7.3 (Muthén & Muthén, 2012). Robust maximum likelihood (MLR) extraction was used. For MLR extraction, the fit indices reported by M*plus* include MLR-based model fit chi-square, the root mean square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and standardized root mean square residual (SRMR). At the statistical levels, model fit is examined using chi-square, with a nonsignificant values indicating good fit. However as chi-square values are inflated with large samples sizes, the RMSEA, CFI, TLI and SRMR values (generally referred to as practical or approximate fit indices) are used. Of the approximate fit indices reported in M*plus*, Hu and Bentler (1998) have recommended a two-index approach to evaluating model fit that includes the SRMR and either the TLI, or CFI, or RMSEA. We used this recommendation to evaluate model fit in the current study. According to the widely used and cited guidelines proposed by Hu and Bentler (1999), RMSEA ≤ .06, CFI and TLI ≥ .95, and SRMR ≤ .08 indicate cutoff levels for good model fit. Relatedly, values of CFI between .90 and .95, RMSEA between .06 and .08, and SRMR .08 and .10 indicate adequate model fit. Although the differences in chi-square values can be used to compare nested models, these values are also inflated by large samples sizes. Therefore, we used differences involving the approximate fit indices. For this, the recommendation is that difference for the RMSEA of ≥ 0.015, and for the CFI of ≥ -.010 can be interpreted as lack of invariance (Chen, 2007; Cheung & Rensvold, 2002). However, as the RMSEA corrects for parsimony, and the CFI does not, it has been proposed that more emphasis be placed on the RMSEA when comparing models estimated with ESEM as more parameters are estimated using this approach (Marsh, 2007; Marsh et al., 2013). We also followed this recommendation in the current paper, thereby placing more emphasis on the RMSEA than CFI values.

Results

Missing Values

There were no missing values in the data set used. No excluded cases either?!

Fit for the Six-Factor ESEM RST-PQ Model for Gender and Age Groups

Prior to testing measurement invariance, the fit of the six-factor ESEM RST-PQ model was examined for ratings provided by males, females, emerging adults and adults. The findings are shown in Table 1. As shown, although the CFI and TLI values indicted poor fit for the six-factor ESEM RST-PQ model for all groups, the RMSEA and SMSR values indicated good fit for all groups. As the two-fit indices approach recommended by Hu and Bentler (1998) proposes that a model with good SRMR and the RMSEA values can be interpreted as indicative of good model-data fit, we interpreted our finding as supportive for the six-factor ESEM RST-PQ model for all the four study groups (males, females, emerging adults and adults) examined.

Measurement Invariance Across Males and Females, and Emerging Adult and Adults

Multiple-Group ESEM Analyses. The results for testing measurement invariance across males and females ratings on the items in the RST-PQ are shown in Table 2. The results across emerging adults and adult are shown in Table 3. Given that we inferred good model-data fit in the single group analyses on the basis of the RMSEA values (but not the CFI values), the difference between the different invariance models was evaluated on the basis of difference in only the RMSEA value. As shown in Tables 2 and 3 for both comparisons, the fit indices for configural invariance (M1 in Tables 2 and 3) indicate good fit in terms of RMSEA and SRMR values, thereby providing support for the configural invariance models. As also shown in Table 2 and3, based on the cut-off score recommended for the RMSEA to infer lack of measurement invariance (≥ 0.015, Chen, 2007; Cheung & Rensvold, 2002), there was no difference between the respective configural invariance models and the metric invariance models (M2 in Tables 2 and 3), the metric invariance models and the scalar invariance models (M3 in Tables 2 and3), and the scalar invariance models and the residual invariance models (M4 in Table 2 and 3). Taken together, these findings indicate full support for the measurement model in terms of metric invariance, scalar invariance, and residual invariance across males and females, and across emerging adults and adults. Tables 2 and 3 also show that there was no difference for the RMSEA values between the residual invariance model and the variances-covariances invariance model (M5 in Tables 2 and 3), thereby indicating support for equivalency for all factor variances and covariances. Also, there was no difference between the variances-covariances invariance model and the latent factor means invariance model (M6 in Tables 2 and 3), thereby indicating support for equivalency for all factor mean scores. Thus there was support for full structural invariance across males and females, and across emerging adults and adults.

MIMIC Analyses. The results of the MIMIC analyses with gender as the predictor, and age as the predictor are both presented in Table 4. For both comparisons, the fit values of the baseline models with all paths from the items to the relevant predictors constrained to zero showed good fit in terms of their RMSEA and SRMR values. For both comparisons, item 45 showed the highest modification index values. The RMSEA values of a MIMIC model in which the paths from this item to the respective predictors freely estimated on the original baseline MIMC models were below the cut-off ($\geqslant 0.015$) for inferring difference in model fit. These findings indicate that none of the items for both comparisons showed DIF, thereby indicating support for invariance across gender even when age was controlled; and across age even when gender was controlled. As shown in Table 4, for both comparisons, the RMSEA values of a revised MIMIC model in which all paths from the latent factors to the relevant predictors were fixed to zero also did not differ from the original MIMIC model, thereby indicating that none of the latent factors differed across males and females when age was controlled; and across age when gender was controlled.

Discussion

The findings in multiple-group ESEM analyses indicated support for the configural model (same pattern of factor structure), and for full measurement invariance for the metric (same factor loadings), scalar (same observed score for the same level latent score), and residual invariances (same unique variances) models for ratings across males and female adults, and emerging adults and adults. For both comparisons, the findings showed invariance for all factor variances and covariances, thereby supporting structural invariance for ratings across males and female adults, and emerging adults and adults. In the ESEM MIMIC analyses that examined DIF as a function of gender, controlling for age, none of the items showed DIF. Also, in the ESEM MIMIC analyses that examined DIF as a function of age, controlling for gender, none of the items showed DIF. Additionally, the findings in both multiple-group ESEM and the ESEM MIMIC analyses showed no difference in all six latent mean scores across males and females; and emerging adults and adults.

Although our findings generally concur with existing data, they also extend existing them. First, consistent with existing data we found support for the six-factor ESEM model in four different groups, thereby attesting to the robustness of this model. Second, the findings for measurement invariance across gender and age is somewhat consistent with exiting data (Eriksson et al., 2019; Pugnaghi et al., 2018). The study by Eriksson et al. that used multiple-group CFA to examine invariance across young adults (< 44 years) and older adults (> 44 years); and females and males for a revised version of the RST-PQ (with 52 items rather than the original 65 items) also reported support for configural, and full metric and residual invariance. However, unlike our findings, they found lack of scalar invariance for six items when younger and older adults were compared, and one item when males and females were compared. Using the MIMIC approach, for the theorized six-factor CFA model, Pugnaghi et al found support for full measurement invariance across gender groups and for age. Additionally, our findings are consistent with the conclusions made in the meta-analysis study by Dong and Dumas (2020) that generally there is invariance across age and gender for personality measures. Despite the comparability of our findings with existing data, our findings extend existing data as our findings examined and supported equivalency for structural invariance and latent mean scores across gender and age. Also, we tested invariance for the six-factor ESEM model that has been proposed as the preferred RST-PQ model (Gomez et al., 2020).

The findings in this study have important implications for the use of the RST-PQ. First, our findings indicated robust support for the six-factor ESEM model for the RST-PQ is consistent with the theoretically proposed six factors in this measure. Second, the support of full measurement invariance, and the absence of DIF items in the MIMIC analyses mean that the ratings of all the items in RST-PQ can be used for comparing male and female adults regardless of age, and younger and older age groups, regardless of gender. Related to this, our findings suggest that for adults, the same normative scores could be used regardless of age or gender. Third, the support for full metric invariance indicated that it is meaningful to compare the males and females, and emerging adults and adults groups for factor correlations and correlations of the factors with other constructs. Also, the support for full scalar invariance indicated that it is meaningful to compare the mean latent scores of these groups, and the support for full residual invariance indicates that it is meaning to compare the mean observed scores across these groups.

In concluding, the study needs to be viewed with some limitations in mind. First, as the findings are based on a single study, there is need for cross-validation of the findings before they can be generalized. Second, since ethnicity was not controlled in the current study, and as this can potentially influence ratings of personality questionnaires (Dong & Dumas, 2020), it is conceivable that our findings are confounded. What about possible cross-cultural differences? Even different language RST-PQ versions? Third, all the participants in this study were from the general community and were not selected randomly. These many have further confounded the generalizability of our findings. Notwithstanding these limitations, the findings in this study and previous psychometric studies of the RST-PQ do indeed provide strong support for the use of the RST-PQ in studies relevant to r-RST. It will be useful for future studies to conduct more studies in this area, keeping in mind the limitations mentioned here.

References

Arnett, J. J. (2004). *Emerging adulthood: The winding road from late teens through the twenties*. Oxford, England: Oxford University Press

Bacon, A. M., Corr, P. J., & Satchell, L. P. (2018). A reinforcement sensitivity theory explanation of antisocial behaviour. *Personality and Individual Differences, 123*, 87–93.

Beaton, A. A., Mutinelli, S., & Corr, P. J. (2017). Fractionating negative and positive affectivity in handedness: Insights from the reinforcement sensitivity theory of personality. *Laterality*, 22(4), 419–444.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Corr, P. J., & Cooper, A. J. (2016). The Reinforcement Sensitivity Theory of Personality Questionnaire (RST-PQ): Development and validation. *Psychological Assessment*, *28*(11), 1427–1400.

Corr, P. J., & McNaughton, N. (2012). Neuroscience and approach/avoidance personality traits: A two stage (valuation-motivation) approach. *Neuroscience and Biobehavioral Reviews*, *36*(10), 2339–2354.

Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences*, *160*, 1-23.

Eriksson, L. J. K., Jansson, B., & Sundin, Ö. Ö. (2019). Psychometric properties of a Swedish version of the reinforcement sensitivity theory of personality questionnaire. *Nordic Psychology*, *71*(2), 134–145.

Gomez, R., & Vance, A. (2008). Parent ratings of ADHD symptoms: Differential symptom functioning across Malaysian Malay and Chinese children. *Journal of Abnormal Child Psychology, 36,* 955-967.

Gomez, R., Watson, S., Van Wynen, J., Trawley, S., Stavropoulos, V., & Corr, P. J (2020). Reinforcement Sensitivity Theory of Personality Questionnaire: Factor structure based on CFA and ESEM, and associations with ADHD. *Journal of Personality Assessment*. On line

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Jiang, Y., & Tiliopoulos, N. (2014). Individual differences in adult attachment and reinforcement sensitivity. Personality and Individual Differences, *68*, 205–210.

Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association, 70*, 631-639.

Krupić, D., Corr, P. J., Ručević, S., Križanić, V., & Gračanin, A. (2016). Five reinforcement sensitivity theory (RST) of personality questionnaires: Comparison, validity and generalization. *Personality and Individual Differences*, *97*, 19–24.

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

Muthen, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer, & H. Braun (Eds.), *Test validity* (pp. 213-238). Hillsdale, NJ: Erlbaum Associates.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables; user's guide [version 7]*.

Pugnaghi, G., Cooper, A., Ettinger, U., & Corr, P. J. (2018). The psychometric properties of the German language reinforcement sensitivity theory-personality questionnaire (RST-PQ). *Journal of Individual Differences*, *39*(3), 182–190.

Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.

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

Wytykowska, A., Fajkowska, M., Domaradzka, E., & Jankowski, K. S. (2017). Construct validity of the Polish version of the reinforcement sensitivity theory-personality questionnaire. *Personality and Individual Differences*, *109*, 172–180.

Table 1

Fit of the Six-Factor ESEM RST-PQ Models

Group	*df*	χ^2	RMSEA		CFI	TLI	SRMR	*N*
			Estimate	90% CI				
Males	1705	2928.26	.056	[.053, .059]	.813	.772	.042	229
Females	1705	4312.90	.048	[.046, .049]	.856	.824	.034	672
Emerging adults	1705	3985.91	.048	[.046, .050]	.850	.817	.035	588
Adults	1705	378.03	.075	[.065, .085]	.825	.785	.040	313
All	1705	3268.29	.054	[.051, .057]	.860	.829	.032	901

Note. χ^2= maximum likelihood χ^2, RMSEA= root mean square error of approximation; CFI= comparative fit index; TLI = Tucker Lewis index; SRMR = standardized root mean square residual; ESEM = exploratory structural equation modeling; *N* = number of participants tested. All χ^2 values were significant ($p < .01$).

Table 2

Results of Tests for Invariance Across Males and Females for the RST-PQ for the Six-Factor ESEM Model

	Model Fit						Model Difference (Δ)		
Model (M)	χ^2	df	RMSEA (90% C.I.)	CFI	TLI	SRMR	ΔM	ΔRMSEA	ΔCFI
M1: Configural invariance	7276.97	3410	.050 (,049-.052)	.844	.810	.036	-	-	-
M2: Weak/metric invariance (M1 with all like item loadings constrained equal)	7633.81	3764	.048 (.046-.049)	. 844	.828	.046	M2 – M1	-.002	.000
M3: Strong/Scalar invariance (M2 with all like item intercepts constrained equal)	7846.88	3823	.048 (,047-.050)	. 838	.823	.047	M3 – M2	.000	-.006
M4: Strict/Residals invariance (M3 with all like item error variances constrained equal	8006.96	3888	.048 (,047-.050)	. 834	.822	.050	M4 – M3	.000	-.004
M5 Invariance for Variance-Covariance	8040.80	3909	.048 (.047-.050)	.833	.823	.055	M5 – M4	.000	-.001
M6. Invariance for Latent Mean	8146.79	3915	.049 (.047-.050)	829	.819	.059	M6 – M5	.001	-.004

Note. χ^2= robust maximum likelihood chi-square (MLRχ^2), RMSEA= root mean square error of approximation; CFI= comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual, C.I. = confidence interval All MLRχ^2 values were significant (p < .001).

Table 3

Results of Tests for Invariance Across Emerging Adults (N = 588) and Adults (N = 313) for the RST-PQ for the Six-Factor ESEM Model

	Model Fit						Model Difference (Δ)		
Model (M)	χ^2	df	RMSEA (90% C.I.)	CFI	TLI	SRMR	ΔM	ΔRMSEA	ΔCFI
M1: Configural invariance	7259.14	3410	.050 (,048-.052)	.841	.805	.037	-	-	-
M2: Weak/metric invariance (M1 with all like item loadings constrained equal)	7718.03	3764	.048 (.047-.050)	. 836	.819	.046	M2 – M1	-.002	.000
M3: Strong/Scalar invariance (M2 with all like item intercepts constrained equal)	8099.46	3823	.050 (,048-.051)	. 823	.807	.050	M3 – M2	.002	-.013
M4: Strict/Residuals invariance (M3 with all like item error variances constrained equal	8230.24	3888	.050 (,048-.051)	. 820	.808	.054	M4 – M3	.000	-.003
M5 Invariance for Variance-Covariance	8262.41	3909	.050 (.048-.051)	.820	.808	.061	M5 – M4	.000	.000
M6. Invariance for Latent Mean	8415.84	3915	.051 (.049-.052)	814	.802	.072	M6 – M5	.001	-.006

Note. χ^2= robust maximum likelihood chi-square (MLRχ^2), RMSEA= root mean square error of approximation; CFI= comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual, C.I. = confidence interval. All MLRχ^2 values were significant (p < .001).

Table 4

ESEM MIMIC Results of Tests for DIF Across Males and Females, and Emerging Adults and Adults for the RST-PQ

	Model Fit						Model Difference (Δ)		
Model (M)	χ^2	df	RMSEA (90% C.I.)	CFI	TLI	SRMR	ΔM	ΔRMSEA	ΔCFI
			Predictor = Gender; Covariate = Age						
MM1: Baseline models (all paths between items and predictor set to zero)	5762.00	1829	.049 (,047-.050)	.842	.809	.044	-	-	-
MM2: MM1 with path for 45 item freed)	5737.49	1828	.049 (.047-.050)	. 843	.810	.044	MM2 – MM1	.000	.001
MM3: MM1 with all paths from latent factors to predictor fixed to zero	5870.48	1835	.049 (,048-.051)	. 838	.805	.047	MM3 – MM1	.000	-.004
			Predictor = Age; Covariate = Gender						
MM1: Baseline models (all paths between items and predictor set to zero)	5680.15	1829	.048 (,047-.050)	. 846	.813	.038			
MM2: (MM1 with path for 45 item freed)	5667.81	1828	.048 (.047-.050)	.846	.814	.038	MM2 – MM1	.000	.000
MM3: (MM1 with all paths from latent factors to predictor fixed to zero)	5868.90	1835	.049 (.045-.051)	839	.805	.047	MM3 – MM1	.001	-.007

Note. χ^2= robust maximum likelihood chi-square (MLRχ^2), RMSEA= root mean square error of approximation; CFI= comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual, C.I. = confidence interval All MLRχ^2 values were significant ($p < .001$).