ESTABLISHING MEASUREMENT INVARIANCE OF THE MMPI-2 RESTRCTURED CLINICAL SCALE 4 (ANTISOCIAL BEHAVIOR) USING AMERICAN AND KOREAN CLINICAL SAMPLES

Jiebing Wang

A journal article submitted in partial fulfillment of
the requirements for the degree of
Master of Science

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
October 2014

ACKNOWLEDGEMENTS

There are many individuals who have contributed to this project that I wish to acknowledge. I would like to thank my committee chair, Dr. Kyunghee Han, for her continual encouragement, guidance, and support throughout this project; my committee members, Dr. Nathan Weed and Dr. Stephen Colarelli, for their insightful suggestions and feedback. I would like to thank Dr. Holly Ketterer and other members in my lab for their previous work on the project; Dr. Kyunghee Han and Dr. Yossef Ben-Porath for their datasets. I would also like to express my thanks to Barb Houghton, Robin Decker, and Ruth Stahl for their help and support.

I am grateful for the prayers of my brothers and sisters in my church and Chinese fellowship families, also, my parents for their continual support and prayers. Last and mostly, I would like to thank my Heavenly Father because "'Tis grace has brought me safe thus far, and grace will lead me home"!

Soli Deo Gloria!

ABSTRACT

ESTABLISHING MEASUREMENT INVARIANCE OF THE MMPI-2 RESTRCTURED CLINICAL SCALE 4 (ANTISOCIAL BEHAVIOR) USING AMERICAN AND KOREAN CLINICAL SAMPLES

by Jiebing Wang

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001), one of the most widely used instruments in personality and psychopathology, has been translated and adapted into many other languages (University of Minnesota Press, 2014). It is important to demonstrate that the translated version shows measurement equivalence or invariance before making mean comparisons across cultures. Ketterer (2010) examined the measurement invariance of the MMPI-2 Restructured Clinical (RC) Scales across American and Korean versions using two normative samples and found that the one-factor model showed poor fit in most scales, making further measurement invariance examination less feasible. Expanding on her study, the present study tested the measurement invariance of the RC4 (Antisocial Behavior) rationally-derived four-factor model (RC4 four-factor model; Han, et al., 2011) across the American and Korean clinical samples using multi-group confirmatory factor analysis (MGCFA). After adjusting for age and gender, the partial scalar invariance of the four factors was achieved, with seven non-invariant items being identified. The Korean clinical sample had a lower latent mean on School Problems and Violation of Social Norms than the American clinical sample, but no difference on Substance Abuse and Family Problems factors. The results supported for the generalizability of the RC4 four-factor model across cultures, however, special attention is needed when using the seven non-invariant items among Korean clinical population.

TABLE OF CONTENTS

Introduction

MMPI-2 Restructured Clinical Scales

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) has been widely used in measuring personality and psychopathology in diverse settings. It eliminated some of the major shortcomings of the previous version of the MMPI (biased norms, outdated item content, etc.), but it kept the original Clinical Scales, despite of their psychometric weaknesses. Having been constructed using an empirical keying strategy, the Clinical Scales contained heterogeneous items and were substantially correlated, which made interpretation of the Clinical Scales complex. Tellegen and his colleagues developed the Restructured Clinical Scales (MMPI-2 RC; Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003) to correct these weaknesses. The RC Scales were developed through the following four steps: (a) identifying the "Demoralization" construct which contained items measuring a general distress and appeared in most Clinical Scales; (b) identifying the distinctive core construct of the Clinical Scales besides the Demoralization; (c) constructing seed scales for the Demoralization and the core constructs; and (d) selecting relevant items for the seed scales to derive the RC Scales (Tellegen et al., 2003). The MMPI-2 RC Scales include: RCd (Demoralization), RC1 (Somatic Complaints), RC2 (Low Positive Emotionality), RC3 (Cynicism), RC4 (Antisocial Behavior), RC6 (Ideals of Persecution), RC7 (Dysfunctional Negative Emotions), RC8 (Aberrant Experiences), and RC9 (Hypomanic Activation).

RC4 (Antisocial Behavior)

The RC scales are shorter than the original Clinical Scales. However, they are reported to be more reliable, show weaker intercorrelations between the scales, and have higher convergent

and discriminant validity than the Clinical Scales (Ben-Porath, 2012). For example, the MMPI-2 Clinical Scale 4 (Psychopathic Deviate; Pd) contains 50 items and measures "the difficulties in school and/or with the law, a lack of concern about most social and moral standards of conduct, the presence of family problems, and absence of life satisfaction" (Butcher et al., 2001, p.26). The RC4 includes 22 items and assesses "a broad spectrum of conduct, including adult and/or juvenile criminal behavior, various other manifestations of juvenile misconduct, substance abuse, aggressive behavior, familial conflict, impulsive behavior, and deceit" (Ben-Porath, 2012, p. 70). Nine of these 22 RC4 items belong to its original Clinical Scale 4.

For the Clinical Scale 4, the internal consistency reliability coefficients were .60 for men and .62 for women in the normative sample, and the test-retest reliability coefficients were .79 for men and .69 for women in the normative sample (Butcher et al., 2001). For the RC4 scale, the internal consistency reliability coefficients were ranged from .77 to .83 in the outpatient and inpatient samples, were .76 for men and .73 for women in the normative samples, and the test-retest reliability coefficient was .89 (Tellegen & Ben-Porath, 2008). The findings above demonstrated that the RC4 had a better overall reliability than its original Clinical Scale 4 despite of its shortened length of items.

Sellbom and his colleagues (2007) examined the convergent and discriminant validity of the Clinical Scale 4 and the RC4 by correlating them with Psychopathy Checklist-Screening Version (PCL-SV; Hart, Cox, & Hart, 1995), conceptually relevant clients' historical records (histories of juvenile delinquency, adult criminality, and substance abuse problems), and conceptually non-relevant emotional distress variables in a forensic setting. Compared to the Clinical Scale 4, the RC4 showed stronger correlations with both facets of the PCL-SV (Affective-Interpersonal and Social Deviance) and antisocial behavior variables, and weaker

correlations with emotional distress variables. The results indicated that convergent and discriminant validity was higher for the RC4 than for its original Clinical Scale counterpart in assessing psychopathy and antisociality.

The improved convergent and discriminant validity of the RC4 compared to the Clinical Scale 4 was also found in a college clinical setting. Compared to the Clinical Scale 4, the RC4 correlated stronger with conceptually relevant criteria (therapists' rating of client hostility, clients' past and present alcohol use, and client exhibition of antisocial and aggressive traits) while correlated weaker with other conceptually non-relevant criteria (Sellbom, Ben-Porath, & Graham, 2006). However, in another study comparing the behavioral and clinical correlates of the Clinical Scale 4 and the RC4 in an outpatient clinical setting, the two scales were similar in predicting antisocial behaviors with the exception being that the RC4 has a better prediction of drug abuse history (Binford & Liljequist, 2008). In inpatient clinical settings, the RC4 scores were related with intoxication, substance abuse in admission, and suicidal ideation. Psychiatric patients with the elevated RC4 scores were more likely to have history of substance abuse problems, violent behavior, and contact with the criminal justice system (Arbisi, Sellbom, & Ben-Porath, 2008).

When using the MMPI-2 in selecting police officers, the RC4 was found as one of the best predictors in negative outcomes and sustained supervisor complaints (Sellbom, Fischler, & Ben-Porath, 2007). In a group of men who were convicted of domestic violence, the RC4 was related to their criminal history, juvenile conduct, mental health problems, anger frequency, and amount of partner violence. Individuals with the elevated RC4 scores were more likely to have batterers' intervention program dismissal and recidivism after it (Sellbom, Ben-Porath, Baum, Erex, & Gregory, 2008).

International Adaptation of MMPI-2/MMPI-2-RF and Measurement Invariance Tests

The MMPI-2/MMPI-2-RF has been translated into many languages and adapted into many cultures. Twenty seven non-English MMPI-2/MMPI-2-RF translations have been published as of 2014 (University of Minnesota Press, 2014). Translating and importing established inventory make cross-cultural comparisons of the interested construct possible. The goal of the comparative cross-cultural research is to distinguish between the universal characteristics (generalized to two or more cultures) and the unique ones (specific to one culture), and the majority of research compare the means across cultural groups on psychological constructs using self-report measures. There are several measurement issues that need to be considered before the mean comparisons are made. These issues are illustrated below using antisocial behavior construct as an example.

One of the issues is that the broad antisocial behavior construct may have different underlying facets and/or different weights of those facets across cultures (Dinges, Atlis, & Vincent, 1997). For example, the diagnostic framework of antisocial personality disorder contains three parts: criminal behaviors—behaviors that are deviant from the legal norms (e.g., physical fights); interpersonal antisocial behaviors—behaviors that put an individual into conflicts with others in his/her social environment (e.g., inability to keep intimate relationships); intrapersonal antisocial behaviors—antisocial behaviors that happened internally (e.g., impulsivity). The weight of these three parts in defining antisocial personality disorder may be different in a more egocentric oriented culture (more emphasis on the self and independence) than sociocentric oriented culture (more emphasis on the being harmony with others and interdependence) (Dinges, Atlis, & Vincent, 1997). Another issue is that an antisocial behavior measure may contain items that relate to the cultural norms of behavior appropriateness. For

example, the use of drug as an indicator of antisocial behavior may not be appropriate in the Korean culture as drugs are available but difficult to access to in the Korea (Kwon-Ahn, 2001).

The issue of measurement invariance needs to be considered when making cross-cultural comparisons on mean scales. Measurement invariance refers to the measurement equivalence of the underlying construct in two or more groups (Meredith, 1993). If the assumption of measurement invariance holds, that is, the same construct being assessed across groups, the group mean comparisons are valid and reflect the true differences on the construct. Many statistical techniques have been used to test measurement invariance. The most frequently used technique is multiple-group confirmatory factor analysis (MGCFA). MGCFA uses a series of hierarchical models, and enables researchers to test equivalence of the factor structure, factor loadings, intercepts, residual variances, and other features of the construct in two or more groups (Chen, 2008; Vandenberg & Lance, 2000).

Configural invariance (or construct invariance), the most basic level of invariance, indicates that similar latent constructs are measured across groups. It is established when the pattern of the factors is similar, the same items being associated with the same factor, across groups, but the factor loadings may differ in groups (Horn, McArdle, & Mason, 1983). For example, cofigural invariance was not met for the Center for Epidemiologic Studies Depression Scale (CES-D) across ethnic groups. A meta-analysis of the confirmatory factor analysis (CFA) studies showed that an additional factor—sorrow/grief—was identified among Asian Americans besides the four original conceptualized factors that other ethnic groups shared for the CES-D (depressed affect, positive affect, somatic symptoms, and interpersonal problem) (Kim, DeCoster, Huang, & Chiriboga, 2011).

Metric invariance (or factor loading invariance), the second level of invariance, is achieved when factor loadings are equal across groups. Factor loadings (λs), coefficients obtained by regressing items on their underlying factors, reflect the magnitude of the relationship between the items and their latent factors. When equal factor loadings across groups hold, it assumes that the strength of the relationships between items and their underlying constructs are the same across groups. In other words, indicators measure the latent construct in the same scale interval or measurement unit across groups (Rock, Werts, & Flaugher, 1978).

Scalar invariance (or intercept invariance), the third level of invariance, exists when factor loadings and intercepts are equal across groups. Intercept is the predicted value of the indictor when the latent factor is zero. When equal intercepts across groups hold, it assumes that the scale has the same measurement unit (factor loading) and origin (intercept) across groups (van de Vijver & Poortinga, 1982). An example of metric non-invariance and scalar non-invariance is the measure of temperature in Celsius (^{o}C) and Fahrenheit (^{o}F). When water freezes at the standard pressure, the temperature is $0^{o}C$ in Celsius measure and $32^{o}F$ in Fahrenheit measure; when water boils at the standard pressure, the temperature is $100^{o}C$ and $212^{o}F$. The change in temperature at two points is 100 in Celsius measure and 180 in Fahrenheit measure, indicating that the two measures have different measurement units. When $0^{o}C$ is measured in Fahrenheit, it is $32^{o}F$; when $0^{o}F$ is measured in Celsius, it is $-17.78^{o}C$, indicating that the two measures have different measurement origins. Without converting one measure to another, it is impossible to compare the temperature in two different measures.

Unique invariance, the fourth level of invariance, is established when the unique variances are equal across groups. Unique variance is the combination of specific factor and measurement error variance. When the unique invariance holds, it demonstrates the same level

of precision of the scale across groups. Measurement invariance is most commonly tested at the level of configural, metric, scalar, and unique variance. For more advanced levels (e.g., variance, covariance, factor means), it is difficult to achieve the measurement invariance (Chen, Sousa, & West, 2005).

Identification of RC4 (Antisocial Behavior) Rational Clusters

Ketterer (2010) examined the measurement invariance of the MMPI-2 RC Scales across American and Korean normative samples using MGCFA. A single-factor model for all the RC Scales was chosen because (a) it was the most parsimonious model; (b) scale-item membership was the same across cultures, making cross-cultural comparisons possible. Tests of the single-factor model showed poor fit for most RC Scales, making measurement invariance examination less feasible. Ketterer concluded that it was possible that most RC Scales fit better in a multiple-factor model than a one-factor model. As a post-hoc test, she performed exploratory factor analysis (EFA) in hope to find consistent multi-factor models across cultures, but failed. She suggested that future research focused on identifying underlying clusters within each of the RC Scales based on rational judgment.

Han and colleagues (2011) (unpublished project) set out a project to identify underlying clusters for each of the RC scales using rational judgment. The judges included two psychology faculty members (one of whom is bilingual in English and Korean) and five psychology graduate students. Regarding RC4, identifying rational clusters involved two phases. In Phase I, they initially identified underlying clusters of RC4 items through three steps. In step one of this phase, seven raters independently reviewed RC4 item content and reported their underlying clusters of the items. In step two, they reached initial consensus on the numbers and themes of the clusters based on two criteria: (a) a cluster should have at least three items; (b) clusters that

the majority of raters (at least 4/7) had identified. In step three, they then reviewed the items of those clusters and selected the items that at least half of those raters identified. As the result, seven raters decided upon four clusters: (a) School Problems; (b) Substance Abuse; (c) Family Problems; and (d) a cluster with name undecided. There was an additional cluster that contained miscellaneous items ("residuals"). In Phase II, they reevaluated the initial item selection for RC4 clusters in two steps. In step one, the raters were asked to consider the possibility of (a) removing poorly fitting items from the current clusters, (b) forming new clusters from "residual" items, (c) putting residual items in the previous clusters, or (d) making no changes. In step two, two residual items were added to the fourth cluster, which was then named "Violation of Social Norms." The raters finalized the clusters: (a) School Problems, (b) Substance Abuse, (c) Family Problems, (d) Violation of Social Norms (see Table 1). The residual item was not included in the measurement invariance analysis.

Present Study

The present study expands on Ketterer's (2010) study by examining the measurement invariance of the RC4 rationally-derived four-factor model (abbreviated as "four-factor model") across American and Korean clinical samples using MGCFA. The following hypotheses are proposed.

As the first step to examine measurement invariance, configural invariance of the four-factor model is tested. For each cultural group, the four-factor model is tested against the one-factor model in model fit, the four-factor model is expected to show a good model fit overall and a better model fit than the one-factor model, indicating similar factor-item structures would be obtained for each group. It has been shown that there is a high probability of achieving configural invariance in previous cross-cultural comparisons, with only nine out of 130 studies

being failed (Chen, 2008). Thus, RC4 four-factor model is expected to achieve configural invariance in the present study.

Hypothesis 1: Configural invariance is expected for the four-factor model of RC4, indicating the latent constructs are similarly defined across cultures. In other words, it is expected that for both cultural groups, the four-factor model would show a good fit overall and a better fit than the one-factor model.

Once configural invariance is achieved, factor loading invariance of the four-factor model is tested. In 74% of the reviewed cross-cultural comparisons, it was found that the average loading was higher in its source culture than in imported culture (Chen, 2008). In Ketterer's (2010) study, the average standardized factor loading of the RC4 items was higher for the American normative sample (.52) than for the Korean normative sample (.42), indicating items which were originally constructed using American culture may not be optimally represent the concept of "antisocial behavior" well in Korean culture (see Table 2). The average standardized factor loading of the RC4 four-factors, therefore, is expected to be higher for the American clinical sample than for the Korean clinical sample.

Hypothesis 2A: Higher average standardized factor loading is expected for the American sample than for the Korean sample.

The previous cross-cultural comparisons show a high probability of achievement of configural invariance. However, about 84% of those comparisons have at least 50% of the factor loadings higher in the source culture than in the imported culture, indicating that the probability of establishment for more rigid forms of measurement invariance after configural invariance might be low (Chen, 2008). Even with lack of full measurement invariance, further invariance tests and other analysis (e.g., factor means) may proceed after the establishment of partial

measurement invariance (Byrne, Shavelson, & Muthén, 1989). For example, if the scale lacks full metric invariance, partial metric invariance can be tested by relaxing the invariant constraints of factor loading on some of the items. If the partial metric invariance is reached, partial scalar invariance may be further tested by relaxing invariant constraints of intercepts on some of the items that are different across groups (Steenkamp & Baumgartner, 1998).

With binary data, factor loadings and thresholds are to be constrained at the same step so metric and scalar invariances are tested in one model—scalar model (Muthén & Muthén, 1998-2012). Ketterer (2010) examined the measurement invariance of MMPI-2 RC Scales across American and Korean normative samples; the equivalent one-factor baseline model was only valid for RCd and RC8 in cultures, and further configural invariance and partial scalar invariance were achieved. Therefore, it is expected that partial scalar invariance is established for the RC4 four-factor model across American and Korean clinical samples in the current study.

Hypothesis 2B: Partial scalar invariance is expected for the RC4 four-factor model across cultures.

Method

Participants

American clinical sample. The MMPI-2 American clinical sample (N = 1,980) consisted of: (a) an inpatient sample from HCMC (Hennepin county medical center) and a Veterans Administration Medical Center (VA); (b) an outpatient sample of clients who were seeking services at the Portage Community Mental Health Center (CMHC) in northeast Ohio during the study period between April 1, 1991, and December 31, 1992. The inpatient sample (N = 1,499; 75.7%) consisted of 1,136 men (M_{age} = 43.02, SD = 14.30) and 363 women (M_{age} =

35.13, *SD* = 12.08). The outpatient sample (*N* = 481; 24.3%) consisted of 179 men (M_{age} = 35.61, *SD* = 10.74) and 302 women (M_{age} = 33.37, *SD* = 9.81).

The initial sample was made up of 2,378 inpatient protocols and 1,219 outpatient protocols, but 161 outpatient cases were eliminated due to MMPI-2 invalidity criteria (Cannot Say ≥ 30, VRIN (T-score) ≥ 80, TRIN (T-score) ≥ 80, and Fp (T-score) > 100). Patients in the Korean clinical sample were diagnosed with primary diagnosis of Schizophrenia, Major Depressive Disorder, Bipolar Disorder, Anxiety Disorder, or Somatoform Disorder. In order to be in correspondence with that of the Korean sample, the U.S. clinical protocols with DSM-III-R (or DSM-IV) Axis I primary diagnosis of one of the five disorders (Schizophrenia, Major Depressive Disorder, Bipolar Disorder, Anxiety Disorder, or Somatoform Disorder) were selected for the final sample (*N* = 1,983). Due to the missing values on the age variable, three cases were further deleted from the selected outpatient protocols. The final selected U.S. inpatient and outpatient protocols (*N* = 1,980) were with primary diagnosis of Schizophrenia (*n* = 194; 9.8%), Depressive Disorder (*n* = 1,220; 61.6%), Bipolar Disorder (*n* = 246; 12.4%), Anxiety Disorder (*n* = 301; 15.2%), or Somatoform Disorder (*n* = 19; 1.0%).

The self-reported racial composition was made up of Caucasian American (82.6%), African American (12.5%), or other (4.9%). Most of the participants were single (40.3%), divorced (27.2%), or married (21.1%). Approximately 51.4% of the sample was unemployed, 26.2% was employed (including self-employed and part-time employed), and 5.2% was retired.

Korean clinical sample. The MMPI-2 Korean clinical sample (*N* = 395) was made up of patients from inpatient and outpatient facilities of Samsung National Hospital in Seoul, Korea. The inpatient sample (*N* = 158; 40%) consisted of 64 men (M_{age} = 36.02, *SD* = 14.99) and 94

women (M_{age} = 40.07, *SD* = 15.39). The outpatient sample (*N* = 237; 60%) consisted of 112 men (M_{age} = 34.86, *SD* = 15.06) and 125 women (M_{age} = 41.19, *SD* = 14.97).

The initial sample was consisted of 400 protocols, but 5 cases were deleted based on MMPI-2 invalidity criteria (Cannot Say ≥ 30, VRIN (T-score) ≥ 80, TRIN (T-score) ≥ 80, and Fp (T-score) > 100). All patients were diagnosed by using DSM-IV except five were using Mini International Neuropsychiatric Interview (MINI). Patients were diagnosed with primary diagnosis of Schizophrenia (*n* = 71; 18.0%), Depressive Disorder (*n* = 85; 21.5%), Bipolar Disorder (*n* =89; 22.5%), Anxiety Disorder (*n* = 86; 21.8%), or Somatoform Disorder (*n* = 64; 16.2%).

Most of the participants were either single (49.1%) or married (44.3%). Approximately 25.32% of the sample was housewives, 22.8% was employed (including self-employed and part-time employed), 21.5% was students, and 21.5% was either unemployed, on leave, or retired.

Measures

Minnesota Multiphasic Personality Inventory-2 Restructured Clinical Scale 4 (MMPI-2 RC4). The Minnesota Multiphasic Personality Inventory-2 (Butch, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) is a clinical self-reported personality assessment and consists of 567 true/false items. The RC4 scale consists of 22 items (16 keyed *true* and 6 keyed *false*). For the normative sample, the internal consistency reliability coefficients of the RC4 were .76 for men and .73 for women, and the test-retest reliability coefficient was .89 (Tellegen & Ben-Porath, 2008). For the current clinical sample, the internal consistency reliability coefficient of the RC4 is .81.

Korean Minnesota Multiphasic Personality Inventory-2 Restructured Clinical Scale 4 (MMPI-2 RC4). The Korean MMPI-2 was originally translated and adapted by Han (1996).

The Korean MMPI-2 was standardized using normative sample and published in 2005 (Kim, Han, Lim, Lee, Min, & Moon, 2005). The internal consistency reliability coefficients of the RC4 were .65 for men and .73 for women in the normative sample, and the test-retest reliability coefficient was .81 (Han, Moon, Lee, & Kim, 2011). For the current clinical sample, the internal consistency reliability coefficient of the RC4 is .73 (Han, Moon, Lee, & Kim, 2011).

Statistical Analysis

Data Preparation. All the falsely keyed items (e.g., "I liked school") were reversely recoded, in the way such that all items were keyed the same direction—a "true" endorsement indicates a greater psychopathology. For the RC4 items, "true" response was coded as "1," "false" response was coded as "0." For the American clinical sample, proportion of missing data for individual RC4 item ranged from 0% to .3%; 35 out of total 1,980 cases (1.8%) have at least one missing value. For the Korean clinical sample, proportion of missing data for individual RC4 item ranged from 0% to 1.3%; 11 out of total 395 cases (2.8%) have at least one missing value. For the missing values, pairwise deletion was used for the analysis. The pairwise deletion strategy with WLSMV estimator was efficient in producing unbiased estimates and standard errors for the parameters when used to analyze data with up to 26% of the missing value for variables and 57% of the cases with at least one missing values (Asparouhov & Muthén, 2010).

Model Estimation Methods. A series of confirmatory factor analysis (CFA) was conducted using Mplus 7.2 (Muthén & Muthén, 1998-2012). Due to the binary nature of the items, a weighted least squares mean variance (WLSMV) and Theta parameterization approach, in controlling over parameterization for binary indicators, were used (Muthén & Muthén, 1998-2012). The tetrachoric correlation matrix was used for the estimation.

For binary indicators, the relationship between an observed binary indicator (y = 0 or 1) and its underlying latent factor η is nonlinear. With WLSMV estimator, M*plus* uses probit function to link the relationship between these two in the linear form (Muthén & Muthén, 2009; Wang & Wang, 2012):

$$P(y = 1 \mid \eta) = \Phi(-\tau+\lambda\eta) \quad \text{(equation 1)}$$

$$\Phi^{-1}[P(y = 1 \mid \eta)] = -\tau+\lambda\eta \quad \text{(equation 2)}$$

$$\text{probit}(y = 1 \mid \eta) = -\tau+\lambda\eta \quad \text{(equation 3)}$$

P: probability function; Φ: standard normal cumulative distribution function; Φ^{-1}: inverse of the standard normal cumulative distribution function; τ = threshold; λ = factor loading; η = a latent variable.

Probit is known as the inverse of the standard normal cumulative distribution function (CDF) and probit of a probability produces a z-value on the x axis of the normal standard distribution (e.g., probit [.025] = -1.96, probit [.50] = 0, probit [.975] = 1.96). For an observed binary indicator y, an unobserved normally distributed continuous response indicator (y^*) is assumed underlying it. Item thresholds, linking the relationship between the y and y^*, are the points on the unobserved normal distribution that divide the distribution into different response categories. The number of the thresholds is the number of the response options minus one (Sass, Castro-Villarreal, McWhirter, McWhirter, & Karcher, 2011). For example, in a binary indicator, only one threshold τ is estimated, which indicating the point on y^* where y=1 if the threshold is exceeded (and where y=0 if the threshold is not exceeded): $y^* > \tau$, y = 1; $y^* \leq \tau$, y = 0 (Wang & Wang, 2012). In Equation 1, factor loading λ can be interpreted in the linear form as 1 unit increase in η results in λ units increase in the probit of getting the observed indicator as 1. For

example, for an item, $\lambda = .70$ can be interpreted as 1 unit increase in the latent construct related to the item results in .70 units increase in the probit of getting endorsement of the item ($y = 1$) (Coxe, West, & Aiken, 2013). In the standardized solution, the y^* variances are assumed as 1. If the standardized factor loading of the item load on the factor is .70, then the variance in y^* that is explained by the latent factor, common variance, is 49% ($.70^2$), while the remaining 52% of the variance in y^* is unique variance (Brown, 2006).

Analytic Procedure. First, CFAs are employed to compare the RC4 four-factor model to the one-factor model in model fit for each cultural group. Second, after the equivalent baseline model (four-factor model) is established, MGCFA is conducted to examine the measurement invariance of the RC4 model across two groups. In the MGCFA, two nested models are tested accordingly: configural model (model 1) and scalar model (model 2). In the configural model, the dataset of two groups are examined simultaneously. In the scalar model, equality constraints—in which unstandardized parameters are set to be equal in value across groups—are placed on the parameters. For the binary indicators, the unstandardized factor loadings and thresholds are constrained to be equal in value across groups in tandem. Therefore, metric invariance and scalar invariance are tested in one model—scalar model (Muthén & Muthén, 1998-2012).

The scalar model (model 2) is compared to the configural model (model 1) in model fit, which provides information on measurement invariance of the RC4 Antisocial Behavior latent construct. If the scale shows lack of full scalar invariance, a partial scalar invariance test is conducted by placing the equality constraints of items of each factor one at a time. The partial scalar model (model 2a, 2b, 2c, or 2d) is compared to the configural model (model 1) in model fit, which provides information on measurement invariance of the RC4 Antisocial Behavior

latent construct at sub-factor level. If the measurement invariance is not met for the partial scalar models (model 2a, 2b, 2c, or 2d), further detection for non-invariant items could be conducted. For example, by relaxing the equality constraints of factor loading and threshold of one item at a time within model 2a, but keeping equally constraining other items of School Problems, a non-invariant item would result in significant model fit difference relative to the scalar model (model 2a).

Evaluation of Model Fit. Chi-square statistic (χ^2), comparative fit index (CFI; with acceptable values above .90, preferable above .95), Tucker-Lewis index (TLI; with acceptable values above .90, preferable above .95), root mean square error of approximation (RMSEA; values below .08 indicate adequate fit, values below .05 indicate close fit) are commonly used fit indices to evaluate the model fit (Bentler, 1990; Brown & Cudek, 1993; Hu & Bentler, 1999). These criteria are also valid for the binary data analysis with WLSMV estimator (Yu, 2002). χ^2 statistic is strict for the model fit evaluation and influenced largely by the sample size. Based on χ^2 statistics, the model with large sample size would always be rejected, even if the difference between the observed and predicted variance-covariance matrix is insignificant (Brown, 2006). Therefore, the model fit evaluation for the individual CFA model relies more heavily on CFI, TLI, and RMSEA in this study.

To compare two nested models, a chi-square difference test, $\Delta\chi^2$, is used. For the continuous indicators with maximum likelihood (ML) estimator, it is calculated as the difference of χ^2 value between the two nested models, because the difference in χ^2 value is still in the χ^2 distribution. Changes in CFI (ΔCFI $\leq$ .01), supplemented by RMSEA (ΔRMSEA $\leq$ .015) in two nested models can also be used as indices for multi-group modeling invariance (Chen, 2007). However, for the binary data with WLSMV estimator, the DIFFTEST is used instead of regular

chi-square difference test because the difference in χ^2 value is not distributed as χ^2 (Brown, 2006). Changes in CFA and RMSEA cannot be used as criteria for modeling invariance evaluation in this study, for no related studies have been done with WLSMV estimator.

In all CFA analyses above, a multiple indicators multiple causes (MIMIC) model is used to control covariates—age, gender (see Figure 1). Covariates are variables that associate with both latent variable and their indicators, but they are not the direct measure of the latent variable. Exclusion of the covariates in the analysis may bias the relationship between the latent variable and its indicators. Demographic variables (e.g., gender, age, and ethnicity) are often included as the covariates in the research (Bollen & Bauldry, 2011). In this study, age and gender are treated as covariates, assuming omission of them in the analyses might bias the relations between "antisocial behavior" construct and its items.

Results

Descriptive Statistics

Table 3 presents the percent of endorsement in the keyed direction on RC4 items for each cultural group. A high value indicates a high level of psychopathology. For majority of the items (19 out of 22), the American clinical sample had a higher or equal percent of endorsing "true" compared to the Korean clinical group. Table 4 presents the tetrachoric correlation matrix of the RC4 items by cultures.

Confirmatory Factor Analysis

For each cultural group, a series of CFA was conducted to compare the RC4 one-factor model with the four-factor model in model fit to determine the baseline model (or configural model) for the following measurement invariance tests. Table 5 shows the model fit statistics. Given the large sample size, the chi-square statistics was significant for all models, $p < .001$. For

the American sample, the one-factor model yielded a poor fit to the data based on CFI and TLI fit indices, but an adequate fit based on RMSEA value, while the four-factor model showed improved model fit statistics. For the Korean sample, the one-factor model had a poor fit based on CFI and TLI fit indices, but a close fit based on RMSEA value; while the four-factor model also showed substantial improved model fit statistics, indicating a good-fitting model to the data. For both cultural groups, the four-factor model provided better fit indices than the one-factor model.

Standardized factor loadings and thresholds of the two models are presented in Table 6. For each cultural group, all items loaded more strongly in the four-factor model than in the one-factor model, though the difference was small for most of the items. The results supported the RC4 four-factor model being a better model to the data than the one-factor model. For 12 of the 21 items (57.1%) in the four-factor model, the factor loading was higher or equal for the American clinical group than the Korean clinical group. The average factor loading was slightly higher in the American clinical group (λ_{mean} = .65) than in the Korean clinical group (λ_{mean} = .61). The average threshold was -.61 for Americans and .13 for Koreans, indicating that the American clinical group was more likely to endorse "true" than the Korean clinical group. For both cultural groups, age was negatively associated with each of the four constructs, as well as with one general construct. For both cultural groups, men showed more elevated means on the constructs except Family Problems where women scored slightly higher than men (Table 7).

Measurement Invariance Tests

Measurement invariance tests were conducted to examine whether the underlying construct Antisocial Behavior is equivalent across two groups, and the results are presented in Table 8. First, the configural model (model 1) was tested in which both factor loadings and thresholds

were freely estimated. The model met the criteria for the acceptable model. Next, the scalar model (model 2), in which both factor loadings and thresholds were equally constrained across two groups, was tested. The fit between the current model relative to the configural model (model 1) was evaluated using *p* value of DIFFTEST. There results did not support the full scalar invariance as the difference in fit was significant (p =.001). Thirdly, the partial scalar models (model 2a, 2b, 2c, and 2d), in which both factor loadings and thresholds of all items of each factor were constrained to be equal across two groups at a time while freeing those of items of other factors, were conducted and *p* value of DIFFTEST was used to evaluate the difference in its model fit compared to the configural model (model 1). The results supported the partial scalar invariance of the model 2c at α level of .01, indicating the scalar invariance of the Family Problems factor.

Further effort to detect the non-invariant items for the School Problems (model 2a), Substance Abuse (model 2b) and Violation of Social Norms (model 2d) factors was made. For the School Problems factor, the factor loading and threshold of item 223 being freely estimated ("suspended from school for bad behavior") in model 2a resulted in significant *p* value of DIFFTEST (p = .001). For the Substance Abuse factor, the factor loading and threshold of item 237 ("never take drugs or sleeping pills"), item 49 ("enjoyed using marijuana"), or item 297 ("get high or drunk") being freely estimated in model 2b resulted in significant *p* value of DIFFTEST, respectively (p < .001 for item 237; p < .001 for item 49; p < .001 for item 297). For the Violation of Social Norms factor, the factor loading and threshold of item 38 ("never in trouble of sex behavior"), item 190 ("never in trouble with the law"), or item 21 ("stole things when young") being freely estimated in model 2d resulted in significant *p* value of DIFFTEST, respectively (p = .006 for item 38; p < .001 for item 190; p = .004 for item 21).

Finally, after freeing the equality constraints of factor loadings and thresholds of items 223, 237, 49, 297, 38, 190, and 21, but keeping equally constraining other items in the scalar model (model 2), latent means were compared for four factors across cultures (Table 9). In the model, the American clinical group was the reference group, which latent means for four factors were set as 0. The Korean clinical group had a lower latent mean on all four factors than American clinical group, and the School Problems and Violation of Social Norms factors were significantly different ($p < .01$) but not the Substance Abuse and Family Problems factors.

Discussion

While the MMPI-2 has been widely translated and adapted into different languages, few researchers have investigated its measurement invariance across cultures. Ketterer (2010) examined the measurement invariance of the MMPI-2 Restructured Clinical (RC) Scales across American and Korean normative samples; the one-factor model showed poor fit in most scales, making further measurement invariance tests impossible. Expanding on Ketterer's study, In solving the problem, the present study examined the measurement invariance of the RC4 (Antisocial Behavior) by testing four-factor model against one-factor model, across the American and Korean clinical samples.

The CFA results showed that the RC4 one-factor model fit poorly to the data, which was consistent with the previous study of testing one-factor model using American and Korean normative samples (Ketterer, 2010). However, the RC4 four-factor model was shown being a good model overall and a better model than the one-factor model across cultures, suggesting intercorrelations among RC4 items were explained well by four intercorrelated latent constructs (School Problems, Substance Abuse, Family Problems, and Violation of Social Norms), which supported the hypothesis 2. It was also found that American clinical group had a slightly higher

average standardized factor loading for items than Korean clinical group in the four-factor model, supporting hypothesis 2A.

The measurement invariance of the four-factor model was examined across the American and Korean clinical samples. The results demonstrated the partial scalar invariance of RC4 sub-factors and lack of invariance of seven items (items 223, 237, 49, 297, 38, 190, and 21), which supports hypothesis 2B. This suggested a good generalizability of the School Problems, Substance Abuse, Family Problems, and Violation of Social Norms items across American and Korean samples, regardless of age and gender.

As the results showed partial scalar invariance, a prerequisite for the further cross-cultural comparisons, mean comparisons across two national samples were made on the four latent factors. Substance Abuse and Family Problems were similar, while School Problems and Violation of Social Norms were different between American and Korean samples. However, before interpreting these sub-constructs cross-culturally, special attention needed to be paid to the non-invariant items—individuals with the same latent mean on the constructs, from different cultural groups, responded differently to the items—as well as the possible explanations for the bias before applying non-invariant items cross-culturally.

For the School Problems factor, the Korean sample yielded a lower endorsement on item 223 ("school suspension") than the American sample at the same levels of the School Problems latent score. One of the possibilities was that most Korean students have not been suspended during their school years (Shin, & Koh, 2005). This might indicate that a large number of Korean participants score at or near the low limit of this item, indicating a "floor effect" (Hessling, Schmidt, & Traxel, 2004). After freeing the equality constraints of item 223 across cultures, the Korean sample had a lower latent means on School Problems than the American

sample, which was consistent with the previous literature. In the PISA (Program for International Student Assessment)—an international comparative survey of student achievement across 65 countries, Korean students reported having a higher classroom disciplinary climate score than the U.S. students (Organization for Economic Cooperation and Development [OECD] 2009, Figure IV.4.2).

For the Substance Abuse factor, compared to the American sample, the Korean sample also yielded a lower endorsement on items 237, 49, and 297 at the same levels of the Substance Abuse latent score. The Korean group's under-endorsement on item 237 ("never used prescription drugs"; the item is reversely coded, which a "false" endorsement indicates a greater psychopathology) and item 49 ("like using marijuana") may be related to the restriction of the drug use in the Korea. The laws prohibiting the usage and possession of drugs, including marijuana, were established by the Southern Korean government in 1946 (Kwon-Ahn, 2001). The drugs are available but difficult to access to in the Korea. The under-endorsement on item 297 ("get drunk frequently") could also be interpreted as a higher level of alcohol consumption would be required for endorsing the "true" response for the Korean group. This may be explained by the great tolerance of heavy drinking towards men in the Korean culture. Korean drinking culture encourages men to drink, and drinking together is viewed as one way to share and increase friendships (Kwon-Ahn, 2001). After freeing the equality constraints of these three non-invariant items across cultures, the Korean and American samples had a similar latent means on School Problems factor. However, the interpretation of the latent mean comparison needs to be cautious as three out of the five items of this factor were detected with item bias.

For the Family Problems factor, the scalar invariance of the factor and similar latent means between two cultural samples were found. Because of the effect of globalization, the Korean

family values and structures have changed greatly in the past decades, which may makes cultures difference less prominent on Family Problems. More traditional extended families are replaced by nuclear families, the average size of the family have changed from 5.6 in 1960 to 3.1 in 2000, and the marriage rate is decreasing while the divorce and remarriage rate is increasing (Park, Kim, & Ko, 2000; Yang, 2003).

For the Violation of Social Norms factor, the Korean sample had a lower endorsement on three items than the American sample at the same levels of the factor latent score. The Korean group's under-endorsement on item 38 ("never having problems because of sex behavior"; the item is reversely coded) and 190 ("never having problems with the law"; the item is reversely coded) was possible that a "floor effect" exists in this item for the Korean group as the Korean culture is defined as being a tight culture with stronger social norms and lower tolerance of violations of the norms than the American culture (Gelfand et al., 2011). Compared to Americans, Koreans have more constraints on everyday situations and are higher in self-regulating (Gelfand et al., 2011). The Korean group's under-endorsement on item 21 ("stole things at a young age") may be related to the expectation of proper behaviors that Koreans been given at a young age. An ill-behaved child causes their parents to be criticized for not teaching their child properly, while a well-behaved child brings honor to the family (Yang, 2009). The finding of the Korean sample with a lower factor latent mean compared to the American sample was consistent with the previous literature. In a comparison of thirty-three countries on the tightness of the culture, Korea was ranked as the fifth tightest country, after the Pakistan, Malaysia, India, and Singapore; while Unites States was ranked as the twenty-forth. Compared to the United States, Korea is likely to have more regulations, laws, police per capita, and stricter punishments in general, as well as a better maintaining of social control by the criminal justice

system shown by the lower crime rates. In addition, Korea with a tight culture put more constraints on everyday situations than the Unites States, and Koreans were higher in self-regulating and having a need for structure (Gelfand, et al., 2011).

Two limitations of this study should be noted. First, the cross-cultural comparison in this study may introduce the sample bias as the American outpatient sample was collected between 1991 and 1992 whereas the Korean clinical sample was collected recently. It is possible that the School Problems, Substance Abuse, Family Problems, and Violation of Social Norms measured in the American clinical sample twenty years ago may be different from the constructs measured more recently. The future study should be replicated using more recent data. Second, other important characteristics of patients (e.g., socioeconomic status) that were related to Antisocial Behavior were not considered as covariates in the study, due to the inconsistent background information variables contained by the three datasets.

Despite of these limitations, the present study has several strengths and implications. First, this study may be the first one examining measurement invariance of MMPI-2 RC4 across cultures using clinical datasets. As the MMPI-2 is intended to assess the personality and psychopathology in the adult clinical population, using American and Korean adult clinical samples may add information to the previous study which used American and Korean adult normative samples (Ketterer, 2010). Findings of this study indicate that the Antisocial Behavior construct was similar across American and Korean clinical sample, in general. This further implies that American and Korean cross-cultural comparisons on MMPI-2 RC4 could be done with small risk of cultural bias. This may be due to decreasing distinctions between cultures because of globalization. However, seven biased items should be used carefully in the Korean group. Second, as MMPI-2 RC4 was originally developed in the United States and not for the

cross-cultural applications, some of the symptoms of equal or greater weight from Korean cultural perspective may be ignored in the Antisocial Behavior construct that MMPI-2 RC4 measures. A combined emic-etic approach may be needed for researchers developing a culture-specific Antisocial Behavior scale (Cheung, van de Vijver, & Leong, 2011). Third, this study tested the model with age and gender as covariates, which were found related to the antisocial behavior (Heitzeg, 2011; Lydecker, 2010). Omission of the age and gender in the analyses that bias the relations between the Antisocial Behavior construct and its items could add to the precision of the model.

In conclusion, it is appropriate to use the MMPI-2 RC4 to measure Antisocial Behavior in the Korean clinical settings. Researchers need to pay special attention to the non-invariant items (items 223, 237, 49, 297, 38, 190, and 21) when interpreting mean scores across cultures. In addition, the Korean clinical sample is more likely to have a lower mean score on School Problems and Violation of Social Norms sub-factors than the American clinical sample. Future research could be replicated using a more recent American clinical dataset and further detect whether the item bias is due to the ARS. A combined emic-etic approach may be also needed for constructing a cultural-specific Antisocial Behavior measure in Korea.

References

Arbisi, P. A., Sellbom, M., & Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in psychiatric inpatients. *Journal of Personality Assessment, 90,* 122-128.

Asparouhov, T., & Muthén, B. (2010). Weighted least square estimation with missing data (Technical Appendix). Retrieved from http://www.statmodel.com/download/GstrucMissingRevision.pdf

Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Binford, A., & Liljequist, L. (2008). Behavioral correlates of selected MMPI-2 clinical, content, and restructured clinical scales. *Journal of Personality Assessment, 90,* 608-614.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods, 16,* 265-284.

Brown, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. S. Long (eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2): Manual for administration, scoring, and interpretation, revised edition*. Minneapolis, MN: University of Minnesota Press.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005-1018.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471-492.

Cheung, F. M., van de Vijver, F. J. R., & Leong, F. T. L. (2011). Towards a new approach to the study of personality in culture. *American Psychologist, 66*, 593-603.

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

Cho, Y. L., & Faulkner, W. R. (1993). Conceptions of alcoholism among Koreans and Americans. *The International Journal of the Addiction, 28*, 681-694.

Coxe, S., West, S. G., & Aiken, L. S. (2013). Generalized linear models. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis* (p.26-51). New York, NY: Oxford University Press.

Dinges, N. G., Atlis, M. M., & Vincent, G. M. (1997). Cross-cultural perspectives on antisocial behaviors. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (p. 462-473). New York: John Wiley & Sons.

Forbey, J. D., Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in a nonclinical setting. *Journal of Personality Assessment, 90,* 136-141.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C.,…Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332,* 1100-1104.

Han, K. (1996). The Korean MMPI-2. In J. N. Butcher (Ed.), *International adaptations of MMPI-2: A handbook of research and applications* (pp.88-136). Minneapolis, MN: University of Minnesota Press.

Han, K., Lim, J., Min, B., & Moon, K. (2011). The *Korean MMPI-2-RF user manual*. Seoul, Korea: Maumsarang Inc.

Han, K., Moon, K., Lee, J., & Kim, J. (2011). *The Korean MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restrucutred Form) user manual*. Seoul, Korea: Maumsarang Inc.

Han, K., Weed, N. C., McCabe, B. J., Dykhouse, A. S., McLaughlan, J. K., & Gilson, A. L. (2011). (unpublished project). Identifying rational clusters of MMPI-2 Restructured Clinical (RC) scales.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Manual for the Psychopathy Checklist—Screening Version (PCL-SV).* Toronto, Ontario, Canada: Multi-Health Systems.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis, MN: University of Minnesota Press.

Hawk, S. S., & Perterson, R. A. (1974). Do MMPI psychopathic deviancy scores reflect psychopathic deviancy or just deviancy? *Journal of Personality Assessment, 38*, 362-368.

Heiner, R. (2011). Cross-cultural and historical methodology. In C. Bryant (Ed.), *The Routledge handbook of deviant behavior* (pp.82-88). New York, NY: Routledge.

Heitzeg, N. A. (2011). Differentials in deviance: Race, class, gender, and age. In In C. Bryant (Ed.), *The Routledge handbook of deviant behavior* (pp.53-59). New York, NY: Routledge.

Hessling, R., Schmidt, T., & Traxel, N. (2004). Floor effect. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods.* (pp.391-392). Thousand Oaks, CA: Sage Publications, Inc.

Horn, J. L., McArdle, J. J., Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1*, 179-188.

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

Ketterer, H. L. (2010). Examining the measurement invariance of the MMPI-2 Restructured Clinical (RC) scales across Korean and American normative samples. (Doctoral dissertation). Retrieved from ProQuest. (UMI Number: 3415643)

Kim, G., DeCoster, J., Huang, C., & Chiriboga, D. A. (2011). Race/Ethnicity and the factor structure of the Center for Epidemiologic Studies Depression Scale: A meta-analysis. Cultural Diversity and Ethnic Minority Psychology, 17, 381-396.

Kim, J., Han, K., Lim, J., Lee, J., Min, B., & Moon, K. (2005). The *Korean MMPI-2-RF user manual*. Seoul, Korea: Maumsarang Inc.

Kim, U., & Park, Y. (2000). Confucianism and family values: Their impact on educational achievement in Korea. *Zeitshrift fur Erziehungswissenschaft* (*Journal of Educational Science*), 3, 229-249.

Kwon-Ahn, Y. H. (2001). Substance abuse among Korean Americans: A sociocultural perspective and framework for intervention. In S. L. A. Straussner (Ed.), *Ethnocultural factors in substance abuse treatment* (pp. 418-435). New York, NY: Guilford.

Lydecker, K. P. (2010). Role of gender in the presentation and course of antisocial behavior for alcohol and drug-disordered youth. (Doctoral dissertation). Retrieved from ProQuest. (UMI Number: 3426046)

Marin, G., Gamba, R. J., & Marin, B. V. (1992).Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of cross-cultural psychology, 23*, 498-509.

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

Muthén, L. K., & Muthén, B. O. (2013). *Mplus version 7.1 language addendum*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2009). *Topic 2: Introductory-advanced regression analysis, IRT, factor analysis and structural equation modeling with categorical, censored, and count outcomes* [short course video]. Retrieved from http://www.statmodel.com/course_materials.shtml

Organization for Economic Cooperation and Development. (2009). PISA 2009 results: What makes a school successful? Resources, Policies, and Practices. Retrieved November 11, 2013, from http://www.oecd.org/pisa/pisaproducts/48852721.pdf

Park, Y., Kim, H., & Ko, H. (2002, June). *Household projections for the Republic of Korea*. Paper presented at the 20th Population Census Conference, Ulaanbaatar, Mongolia.

Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across groups. *Multivariate Behavioral Research, 13*, 403-418.

Sass, D. A., Castro-Villarreal, F., McWhirter, B. T., McWhirter, E. H., & Karcher, M. J. (2011). A cross-cultural assessment of school connectedness: Testing measurement invariance with United States and Chilean adolescents. *Journal of Psychoeducational Assessment*, 29, 428-442.

Sellbom, M., Ben-Porath, Y. S., Lilienfeld, S. O., Patrick, C. J., & Graham, J. R. (2005). Assessing psychopathic personality traits with the MMPI-2. *Journal of Personality Assessment, 85*, 334-343.

Sellbom, M., Ben-Porath, Y. S., & Graham, J. R. (2006). Correlates of the MMPI-2 Restructured Clinical (RC) scales in a college counseling setting. *Journal of Personality Assessment, 86*, 89-99.

Sellbom, M., Ben-Porath, Y. S., & Stafford, K. S. (2007). A comparison of MMPI-2 measures of psychopathic deviance in a forensic setting. *Psychological Assessment, 19*, 430-436.

Sellbom, M., Ben-Porath, Y. S., Baum, L. J., Erez, E., & Gregory, C. (2008). Empirical correlates of the MMPI-2 Restructure Clinical (RC) scales in a batterers' interventions program. *Journal of Personality Assessment, 90*, 129-135.

Sellbom, M., Fischler, G. L., & Ben-Porath, Y. S. (2007). Identifying MMPI-2 predictors of police officer integrity and misconduct. *Criminal Justice and Behavior, 34*, 985-1004.

Shin, S., & Koh, M. (2005). Korean education in cultural context. *Essays in Education, 14*. Retrieved from http://www.usca.edu/essays/vol142005/koh.pdf

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78-107.

Tellegen, A., & Ben-Porath, Y. S. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2 RF): Technical manual*. Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 restructured clinical (RC) scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.

University of Minnesota Press (2014). MMPI-2-RF, MMPI-2, and MMPI-A translations. Retrieved October 6, 2014, from http://www.upress.umn.edu/test-division/translations-permissions/permissions.

van de Vijver, F. J. R., & Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology, 13*, 387-408.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practice, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus.* Beijing: Higher Education Press.

Yang, O. K. (2003). Family structure and relations. *Social Indicators Research, 62*, 121-148.

Yang, S. (2009). Cane of love: Parental attitudes towards corporal punishment in Korea. *British Journal of Social Work, 39,* 1540-1555.

Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. (Doctoral dissertation). Retrieved from ProQuest. (UMI Number: 3066425)

Table 1

RC4 (Antisocial Behavior) Rational Clusters

Cluster	Item no.
School Problems	223
	66
	126 (r)
	205
	253
Substance Abuse	141
	237 (r)
	49
	266
	297
Family Problems	5
	80 (r)
	96
	173
	19 (r)
Violation of Social Norms	38 (r)
	218
	190 (r)
	329
	21
	312
Residual	156

Note. Residual item was not included in the measurement invariance analysis. The item numbers are corresponding to MMPI-2-RF items. (r) indicates that the item is reversely coded, which a "false" endorsement indicates a greater psychopathology. The responses of these items were recoded in such way that all items were keyed the same direction.

Table 2

RC4 (Antisocial Behavior) Baseline CFA Factor Loadings, Thresholds, and Model Fit Indices by Culture

		America[1]		Korea[2]	
Item no.[*]		FL	TH	FL	TH
5		.37	.33	.26	-.11
38 (r)		.58	.09	.37	.55
21		.71	1.20	.53	1.22
80 (r)		.72	.87	.56	.75
223		.57	.88	.50	.65
66		.42	1.67	.48	1.26
126 (r)		.55	.46	.53	.69
96		.39	-.24	.08	-.03
218		.38	1.09	.46	.74
141		.63	1.14	.45	.60
190 (r)		.62	1.60	.67	1.27
156		.63	.55	.55	1.46
173		.58	1.65	.61	1.10
205		.71	1.10	.68	.84
237 (r)		.65	1.63	.54	1.20
253		.62	1.16	.61	1.05
19 (r)		.34	.87	.15	.22
49		.27	.79	.26	.19
266		.31	.95	.26	.25
297		.63	.62	.32	.58
312		.61	.60	.19	.30
329		.25	.95	.23	.50
	M	.52	.91	.42	.69
	SD	.15	.50	.18	.45
	Min	.25	-.24	.08	-.11
	Max	.72	1.67	.68	1.46
	χ^2	1284.76		573.49	
	(*df*, *p*)	(151, <.001)		(144, <.001)	
	CFI	.82		.80	
	NNFI	.88		.82	
	RMSEA	.05		.05	

Note. This table was reproduced from Ketterer's Table 8 (2010). (r) indicates that the item is originally reversely coded, where a "false" endorsement is indicative of greater psychopathology. These items were recoded in such way that all items were keyed the same direction. FL = Factor loading. TH = Threshold. CFI = Comparative Fit Index. NNFI = Nonnormed Fit Index. RMSEA = Root Mean Square Error of Approximation. [1]*N* = 2,530. [2]*N* = 1,352. [*] MMPI-2-RF item number. All factor loadings were significant at α = .001

Table 3

Percent of Responding to Keyed Direction on RC4 Items

Clusters	Item no.	America[1]	Korea[2]
School Problems	223	**35**	8
	66	**46**	18
	126 (r)	41	**50**
	205	**39**	31
	253	**24**	9
Substance Abuse	141	**52**	18
	237 (r)	**40**	37
	49	**36**	3
	266	**28**	9
	297	**21**	12
Family Problems	5	**70**	58
	80 (r)	46	**52**
	96	**46**	29
	173	**34**	32
	19 (r)	40	40
Violation of Social Norms	38 (r)	25	**35**
	218	**14**	4
	190 (r)	**59**	22
	329	**28**	14
	21	**63**	34
	312	**24**	11
Residual	156	65	55
	M	39.82	26.41
	SD	15.13	17.28
	Min	14	3
	Max	70	58

Note. [1]*N* = 1,980. [2]*N* = 395. (r) indicates that the item was reversely keyed, which a "false" endorsement indicates a greater psychopathology. These items were recoded in such way that all items were keyed the same direction. For item 126, 41% of American participants reported false on "liked school." Boldface denotes the higher percent endorsement to the keyed direction in one culture over the other.

Table 4

RC4 Item Correlations by Culture

Item no.	223	66	126	205	253	141	237	49	266	297	5	80	96	173	19	38	218	190	329	21	312	156
223		**.68**	**.16**	**.25**	**.49**	.09	.13	.21	.22	.24	-.03	.02	.26	.42	.10	.19	.23	.18	.23	.27	.16	.18
66	**.86**		**.19**	**.41**	**.76**	.41	.26	-.06	.50	.44	.12	.12	.42	.31	.11	.16	.20	.11	.43	.20	.48	.03
126	**.31**	**.35**		**.31**	**.38**	.23	.12	-.22	.23	.06	.09	.21	.17	.23	.21	.06	.07	-.07	.21	.02	-.03	.17
205	**.52**	**.48**	**.44**		**.41**	.15	.10	.21	.29	.16	.14	.25	.32	.33	-.02	.07	-.13	.17	.17	.17	.23	.21
253	**.57**	**.68**	**.45**	**.52**		**.25**	**.30**	**.34**	**.42**	**.31**	.11	.14	.21	.32	.06	.09	.29	.29	.41	.24	.45	.15
141	.35	.33	.14	.26	**.24**		**.28**	**-.20**	**.79**	**.88**	.31	.10	.37	.15	.15	.09	.16	.15	.31	.15	.58	.15
237	.22	.19	.15	.17	**.16**	**.35**		**.26**	**.17**	**.32**	.11	.17	.08	.06	.18	.16	.19	.19	.09	.07	.20	.04
49	.29	.26	.11	.23	**.19**	**.43**	**.53**		**.46**	**.24**	.16	.10	.12	.06	.24	.14	.01	.44	.21	.21	.01	-.13
266	.30	.28	.16	.31	**.26**	**.74**	**.45**	**.44**		**.83**	.24	.28	.46	.15	.13	.22	.03	.35	.46	.25	.74	.13
297	.23	.17	.10	.19	**.14**	**.52**	**.51**	**.44**	**.68**		**.38**	**.19**	**.29**	**.06**	**.18**	.09	.25	.38	.46	.20	.65	.21
5	.22	.26	.19	.28	.20	.22	.22	.22	.26	**.18**		**.18**	**.25**	**.34**	**.21**	.10	.21	.33	.22	.20	.32	.16
80	.17	.21	.09	.12	.23	-.02	.11	.09	.08	**.03**	**.28**		**.21**	**.23**	**.22**	.15	.18	.19	.36	.19	.25	.05
96	.38	.40	.17	.33	.40	.22	.23	.24	.24	**.17**	**.28**	**.21**		**.28**	**.07**	.19	.31	.14	.32	.12	.37	.14
173	.32	.34	.12	.24	.29	.11	.11	.06	.17	**.11**	**.30**	**.26**	**.29**		**.25**	.14	.21	.21	.23	.32	.11	.13
19	.14	.17	.15	.14	.19	.09	.11	.05	.11	**.04**	**.30**	**.45**	**.22**	**.37**		**.07**	**.24**	**.03**	**.02**	**-.01**	**.01**	.03
38	.08	.12	.01	.13	.19	.10	.14	.10	.11	.02	.07	.16	.19	.10	**.11**		**.19**	**.46**	**-.02**	**.21**	**.11**	-.04
218	.21	.29	.07	.23	.26	.13	.18	.19	.18	.17	.27	.17	.26	.18	**.07**	**.20**		**.36**	**.51**	**.56**	**.35**	.20
190	.42	.39	.08	.21	.26	.36	.28	.29	.41	.22	.22	.20	.23	.25	**.09**	**.16**	**.25**		**.35**	**.21**	**.33**	.06
329	.41	.43	.13	.23	.28	.28	.25	.28	.27	.18	.25	.25	.27	.27	**.12**	**.14**	**.30**	**.38**		**.21**	**.61**	-.07
21	.39	.44	.17	.36	.29	.27	.23	.28	.29	.26	.24	.06	.26	.29	**.14**	**.14**	**.45**	**.34**	**.25**		**.18**	.30
312	.31	.30	.13	.29	.25	.57	.29	.28	.54	.33	.22	.22	.28	.25	**.17**	**.12**	**.20**	**.39**	**.51**	**.28**		.30
156	.20	.24	.17	.37	.20	.23	.21	.19	.17	.19	.20	.07	.13	.06	.08	.12	.26	.18	.13	.40	.17	

Note. Values below diagonal are correlation coefficients for American clinical sample (N = 1,980), whereas those above diagonal are correlation coefficients for Korean clinical sample (N = 395). Intercorrelations among items for each factor are in bold and included in a triangular shape.

Table 5

CFA of RC4: Overall Model Fit

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA	90% CI
America[1]							
One-factor model	2982.191	251	<.001	.795	.775	.074	[.072, .077]
Four-factor model	1101.889	217	<.001	.931	.920	.045	[.043, .048]
Korea[2]							
One-factor model	472.336	251	<.001	.851	.836	.047	[.041, .054]
Four-factor model	302.107	217	<.001	.942	.932	.032	[.022, .040]

Notes. [1]*N* = 1,980. [2]*N* = 395. CFI = comparative fit index. TLI = Tucker-Lewis index. RMSEA = root mean square error of approximation. CI = confidence interval.

Table 6

Standardized Factor Loadings and Thresholds of MMPI-2 RC4 (Antisocial Behavior) Models by Culture

		America[1]				Korea[2]			
		One-factor		Four-factor		One-factor		Four-factor	
Cluster	Item no.	FL	TH	FL	TH	FL	TH	FL	TH
School Problems	223	.84	-1.17	.90	-1.18	.53	.66	.64	.66
	66	.87	-1.55	.94	-1.57	.74	-.44	.94	-.44
	126 (r)	.38	-.15	.45	-.15	.31	-.51	.38	-.51
	205	.60	-.37	.66	-.38	.43	.10	.54	.10
	253	.66	-.30	.75	-.30	.68	.68	.82	.68
	M (*SD*)			.74 (.20)	-.72 (.62)			.66 (.22)	.10 (.57)
	r_{FL} (r_{TH})					.83 (.07)			
Substance Abuse	141	.66	-1.16	.80	-1.16	.84	-.27	.88	-.28
	237 (r)	.52	-.78	.64	-.78	.33	.07	.37	.07
	49	.55	-1.39	.68	-1.39	.31	1.10	.34	1.10
	266	.72	-.20	.88	-.20	.88	.58	.93	.59
	297	.57	-.41	.71	-.41	.92	-.05	.96	-.05
	M (*SD*)			.74 (.10)	-.79 (.50)			.70 (.31)	.29 (.56)
	r_{FL} (r_{TH})					.74 (-.21)			
Family Problems	5	.45	-1.35	.62	-1.32	.42	-.40	.53	-.40
	80 (r)	.31	-.43	.48	-.42	.39	-.53	.51	-.52
	96	.52	-.59	.70	-.57	.54	-.33	.66	-.32
	173	.42	.01	.60	.01	.45	-.28	.57	-.28
	19 (r)	.29	.11	.47	.10	.22	.31	.26	.31
	M (*SD*)			.57 (.10)	-.44 (.57)			.51 (.15)	-.24 (.32)
	r_{FL} (r_{TH})					.81 (.58)			
Violation of Social Norms	38 (r)	.22	.50	.24	.50	.29	-.16	.33	-.16
	218	.42	.69	.46	.69	.47	1.04	.55	1.04
	190 (r)	.56	-1.69	.64	-1.68	.44	.63	.51	.63
	329	.57	-.84	.64	-.83	.62	.22	.71	.23
	21	.57	-1.46	.61	-1.46	.43	-1.05	.46	-1.06
	312	.62	-.29	.70	-.29	.74	1.27	.85	1.29
	M (*SD*)			.55 (.17)	-.51 (.99)			.57 (.19)	.33 (.86)
	r_{FL} (r_{TH})					.75 (.37)			
Residual	156	.39	-1.08			.30	-1.50		
	M	.53	-.63	.65	-.61	.51	.05	.61	.13
	SD	.17	.67	.17	.68	.21	.70	.22	.62

Note. (r) indicates that the item is reversely coded, endorsing "false" indicates a greater psychopathology. [1]*N* = 1,980; [2]*N* = 395. FL = Factor loading; all factor. TH = Threshold. *M* (*SD*) = Mean (Standardized Deviation); r_{FL} (r_{TH}) = correlation of factor loadings (thresholds) between the two cultures.

Table 7

Standardized Covariates Estimates

	Age	Gender[3]
America[1]		
One factor model	-.37*	-.23*
Four-factor model		
School Problems	-.30*	-.23*
Substance Abuse	-.29*	-.22*
Family Problems	-.40*	.08*
Violation of Social Norms	-.28*	-.35*
M	-.32	-.18
SD	.06	.18
Korea[2]		
One factor model	-.34*	-.15*
Four-factor model		
School Problems	-.21*	-.19*
Substance Abuse	-.18	-.22*
Family Problems	-.48*	.08
Violation of Social Norms	-.25*	-.11
M	-.28	-.11
SD	.14	.13

Note. [1]*N* = 1,980. [2]*N* = 395. [3] (Men = 0; women = 1). *$p < .01$.

Table 8.

Measurement invariance tests across American and Korean clinical samples

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA	90% CI	Contrast	$\Delta\chi^2$	Δdf	*p*
Model 1	1255.540	434	<.001	.939	.929	.040	[.037, .043]				
Model 2	1338.471	468	<.001	.935	.930	.040	[.037, .042]	2 vs. 1	145.430	34	<.001
Model 2a	1265.146	442	<.001	.939	.930	.040	[.037, .042]	2a vs. 1	27.152	8	.001
223	1253.956	440	<.001	.939	.931	.039	[.037, .042]	223 vs. 2a	13.476	2	.001
66	1267.584	440	<.001	.938	.929	.040	[.037, .042]	66 vs. 2a	.951	2	.622
126	1262.084	440	<.001	.939	.930	.040	[.037, .042]	126 vs. 2a	7.849	2	.020
205	1269.290	440	<.001	.938	.929	.040	[.037, .042]	205 vs. 2a	2.376	2	.305
253	1263.362	440	<.001	.939	.930	.040	[.037, .042]	253 vs. 2a	7.610	2	.022
Model 2b	1331.811	442	<.001	.934	.925	.041	[.039, .044]	2b vs. 1	90.022	8	<.001
141	1329.802	440	<.001	.934	.924	.041	[.039, .044]	141 vs. 2b	7.954	2	.019
237	1292.638	440	<.001	.937	.927	.040	[.038, .043]	237 vs. 2b	33.533	2	<.001
49	1310.110	440	<.001	.935	.926	.041	[.038, .043]	49 vs. 2b	25.987	2	<.001
266	1330.581	440	<.001	.934	.924	.041	[.039. .044]	266 vs. 2b	5.711	2	.058
297	1281.938	440	<.001	.937	.928	.040	[.038, .043]	297 vs. 2b	75.685	2	<.001
Model 2c	1239.075	442	<.001	.941	.932	.039	[.036, .042]	2c vs. 1	13.884	8	.085
Model 2d	1262.888	444	<.001	.939	.931	.039	[.037, .042]	2d vs. 1	35.309	10	<.001
38	1256.654	442	<.001	.939	.931	.039	[.037, .042]	38 vs. 2d	10.064	2	.007
218	1274.615	442	<.001	.938	.929	.040	[.037, .042]	218 vs. 2d	1.890	2	.389
190	1245.521	442	<.001	.940	.932	.039	[.037, .042]	190 vs. 2d	18.116	2	<.001
329	1269.926	442	<.001	.938	.930	.040	[.037, .042]	329 vs. 2d	1.707	2	.426
21	1255.595	442	<.001	.939	.931	.039	[.037, .042]	21 vs. 2d	11.240	2	.004
312	1263.600	442	<.001	.939	.930	.040	[.037, .042]	312 vs. 2d	6.173	2	.046

Notes. N = 1,980 for American clinical sample; N = 395 for Korean clinical sample. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval. Model 1: Configural Model. Model 2: Scalar Model. Model 2a-2d: Partial Scalar Model. In Model 2a (2b, 2c, or 2d), factor loadings and threshold of only the School Problem (Substance Abuse, Family Problems, or Violation of Social Norms) factor items are constrained to be equal across American and Korean clinical samples. The α level of .01 was used.

Table 9

Latent Means of the RC4 Four Factors

	America[1]				Korea[2]			
	Estimate	*SE*	Estimate/*SE*	*p*	Estimate	*SE*	Estimate/*SE*	*p*
School Problems	—	—	—	—	-2.480	.832	-2.981	.003
Substance Abuse	—	—	—	—	-1.385	.648	-2.136	.033
Family Problems	—	—	—	—	-.322	.197	-1.631	.103
Violation of Social Norms	—	—	—	—	-.482	.139	-3.472	.001

Note. [1]*N* = 1,980. [2]*N* = 395. American clinical group, estimates and *SE*s are set as 0, is the reference group which Korean clinical group compared with. The α level of .01 was used.

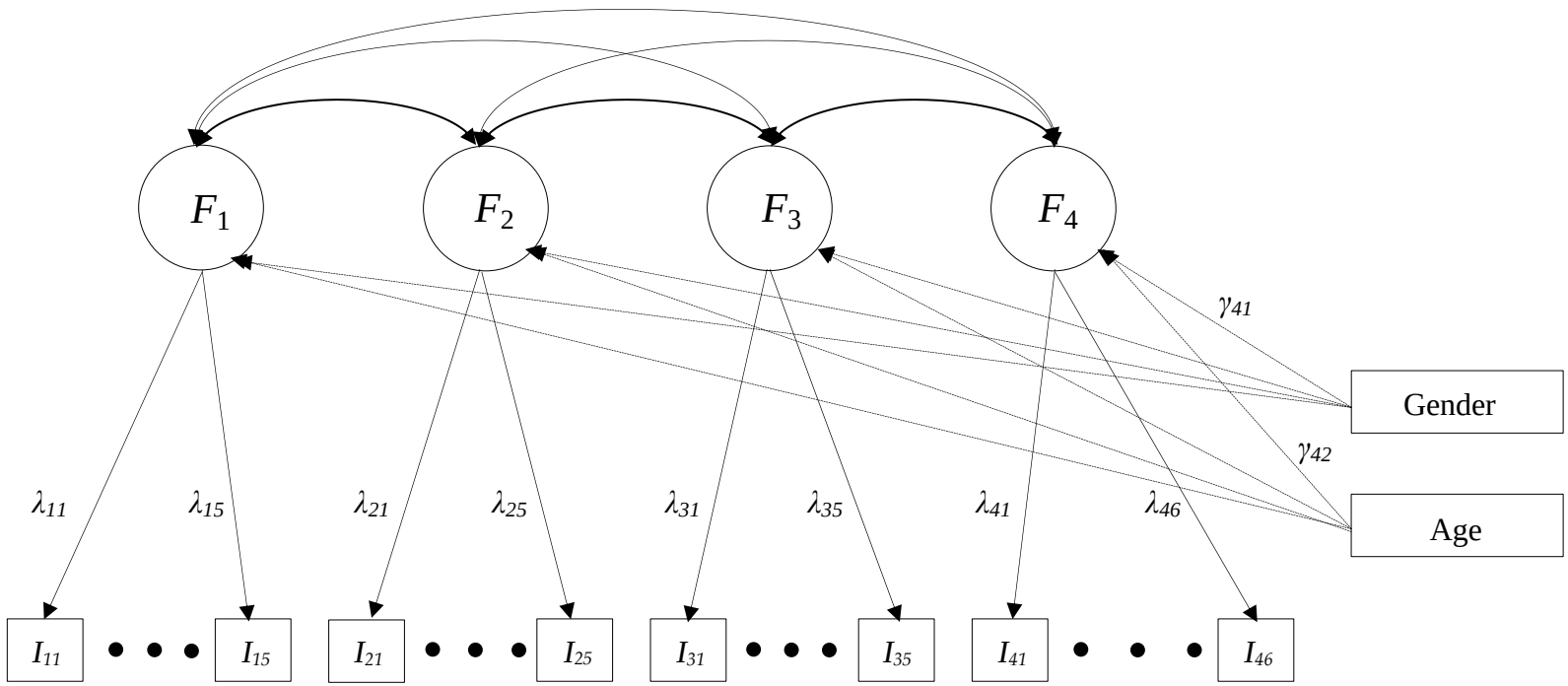

Figure 1. RC4 four-factor model with gender and age as covariates.
Solid lines denote measurement model components; dashed lines denote structural model components. I_{ij} = item *j* of factor *i*; λ_{ij} = factor loading for the *j*th item of factor *i*; γ_{jk} = regression coefficient showing the group mean difference on factor *j* for covariate *k*.