Summer 2018

Examining the Measurement Invariance of the MMPI-A-RF Externalizing Scales Across Korean and American Adolescent Normative Samples

Kristoffer Yong Park
Old Dominion University

EXAMINING THE MEASUREMENT INVARIANCE OF THE MMPI-A-RF EXTERNALIZING SCALES ACROSS KOREAN AND AMERICAN ADOLESCENT NORMATIVE SAMPLES

by

Kristoffer Yong Park
B.S. May 2010, Georgia State University
M.T.S. May 2012, Emory University
M.S. August 2016, Old Dominion University

A Dissertation Submitted to the Graduate Faculties of
Eastern Virginia Medical School
Norfolk State University
Old Dominion University
In Partial Requirement for the Degree of

DOCTORAL OF PHILOSOPHY

CLINICAL PSYCHOLOGY

VIRGINIA CONSORTIUM PROGRAM IN CLINICAL PSYCHOLOGY
August 2018

Approved by:

Jennifer Flaherty (Chair)

Richard Handel (Co-Chair)

Kyunghee Han (Member)

Desideria Hacker (Member)

Barbara Winstead (Member)

ABSTRACT

EXAMINING THE MEASUREMENT INVARIANCE OF THE MMPI-A-RF EXTERNALIZING SCALES ACROSS KOREAN AND AMERICAN ADOLESCENT NORMATIVE SAMPLES

Kristoffer Yong Park
Virginia Consortium Program in Clinical Psychology, 2018
Chair: Dr. Jennifer Flaherty

The Minnesota Multiphasic Personality Inventory – Adolescents – Restructured Form (MMPI-A-RF; Archer, Handel, Ben-Porath, & Tellegen, 2016) is a newly developed instrument in personality and psychopathology and has been translated into many different languages, including the Korean language (University of Minnesota Press, 2016). Due to the cultural differences between Korean and American populations, it is important that constructs measured in the MMPI-A-RF are interpretable across cultures. Focusing on the Externalizing Scales, the present study used multiple-group confirmatory factor analysis (MGCFA) to examine the measurement invariance in Korean and American adolescent normative samples. Partially supporting the hypothesis, the results showed that partial factorial invariance was achieved in four out of the six Externalizing Scales. Noninvariant items of each scale were also identified and cultural differences and implications were discussed. This study expanded the literature on the psychometric properties of the MMPI-A-RF and identified whether any discrepancies between samples reflected true cultural differences, or measurement bias.

ACKNOWLEDGMENTS

I am most grateful for the commitment and endless support of my dissertation chair, Dr. Jennifer Flaherty, my dissertation co-chair, Dr. Richard Handel, and my committee members, Dr. Kyunghee Han, Dr. Desideria Hacker, and Dr. Barbara Winstead. Thank you to the professors, supervisors, friends, and family who have always believed in me.

NOMENCLATURE

Λ	Factor Loadings
Φ	Factor Variances-Covariances
Θ_{δ}	Unique Variances
δ_{COV}	Error Covariance
Φ_{COV}	Factor Covariance
χ^2	Chi-Squared

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

The MMPI-A-RF (Archer et al., 2016) is a multidimensional self-report measure that assesses for aspects of psychopathology and personality functioning in adolescents. It is the second version of the MMPI for adolescent use. Using the MMPI-2-RF as a template, the MMPI-A-RF was developed to address concerns with heterogeneity and intercorrelations in the Clinical scales and to revise and develop the Validity Scales and additional substantive scales. The MMPI-A-RF is a more concise version of the MMPI-A, reducing the total number of items from 478 to 241. The test was validated using normative, inpatient, outpatient, correctional, and school samples to confirm its reliability, and convergent and discriminant validity. In total, there are six Validity Scales, three Higher-Order Scales, nine RC Scales, 25 Specific Problem Scales, and five Personality Psychopathology Scales (Archer et al., 2016). The Specific Problem Scales contain six Externalizing Scales including Negative School Attitudes, Antisocial Attitudes, Conduct Problems, Substance Abuse, Negative Peer Influence, and Aggression. The Externalizing Scales measure components of the Antisocial Behavior and Hypomanic Activation RC Scales.

The family of MMPI assessments has been translated into many different languages including the Korean language (http://www.upress.umn.edu/test-division/translations-permissions/available-translations; Butcher, 1996; Butcher & Pancheri, 1976; Cheung, Lee, & Jin, 1963; Han, 1993; Han, Moon, Lee, & Kim, 2011; Kim et al., 1989). The Korean MMPI-A-RF will be published in the fall of 2018. Due to the differences between American and Korean culture, it is important that constructs measured in the MMPI-A-RF are relevant and interpretable across cultures. The components of the MMPI-A-RF may have varying underlying facets that

can be weighted with different emphasis in the American and Korean population. This study, in particular, focuses on the Externalizing Scales.

Individualism versus collectivism is a cultural concept that distinguishes between Western and Eastern cultures (Le & Stockdale, 2005). How individuals orient their values and behaviors, either internally or within social roles and relationships can play a vital role in illustrating and interpreting one's personality and interpersonal functioning. In the Korean culture, the philosophy of Confucianism contributes to all aspects of society including the home, school, community, and political context (Oh-Hwang, 1993). In addition, although the overall rate of criminal activity in American adolescents has been significantly decreasing since 1996, the delinquency rates among Korean youth have almost doubled in number since 2004 (National Center for Juvenile Justice, 2015; Supreme Prosecutors' Office of Korea, 2011). These cultural and environmental factors may affect the constructs portrayed in the MMPI-A-RF Externalizing Scales and the interpretation of results across the Korean and American adolescent populations.

Although the research examining cultural differences between populations, or test bias, has a variety of approaches, measurement invariance has been the leading method in establishing construct equivalence (Millsap, 1997). Measurement invariance is a concept that ensures that item responses relate to latent constructs in the same way across groups. The present study uses multiple-group confirmatory factor analysis (MGCFA) to examine the measurement invariance in Korean and American samples. Multiple-group confirmatory factor analysis simultaneously estimates model fit between groups and is rooted in CFA, which is a type of structural equation modeling that assesses the relationships between indicators and latent variables. CFA model parameters (factor loadings, factor variances, residual variances, factor means, latent means),

Goodness-of-fit indices, and Modification Indices were analyzed to determine the measurement invariance across cultures (Brown, 2006).

While the research examining measurement invariance in the MMPI-2 Restructured Clinical Scales between Korean and American samples is limited, there are a few studies that have established meaningful results. Ketterer (2010) used multiple-group CFA to examine the measurement invariance of the RC Scales on the MMPI-2 in the Korean and American normative samples. The author found that configural invariance was established for RCd (Demoralization; a general malaise or subjective feelings of distress), RC6 (Ideas of Persecution), and RC8 (Aberrant Experiences) indicating these were similar constructs across groups, while partial factorial invariance was found for RCd and RC6 and full factorial invariance was found for RC8 indicating at least similar meaning and structure of constructs (Tellegan et al., 2003). Wang (2014) also used multiple-group CFA to examine the measurement invariance of the MMPI-2 RC4 (Antisocial Behavior) scale across Korean and American clinical samples. The author found partial scalar invariance for the three latent factors (School Problems, Substance Abuse, Violation of Social Norms) and full scalar invariance for Family Problems demonstrating that individuals with identical scores would experience similar intensity of the construct across cultures. Both authors found results that showed partial consistency in measurement invariance between Korean and American cultures and warranted further studies to establish equivalence between groups.

The purpose of the present study is to examine the measurement invariance of the newly developed MMPI-A-RF Externalizing Scales across Korean and American normative samples. The advantage of using multiple-group CFA is that it allows for the investigation of all aspects of invariance including factor structure, factor loadings, thresholds, residuals, and latent means.

This study has two main objectives. First, since the MMPI literature finds that there are often discrepancies in item scores between Korean and American samples, further testing is needed using the MMPI-A-RF to identify whether these differences account for true cultural differences, or reflections of measurement bias. Second, this study added to the newly developed MMPI-A-RF literature in examining its psychometric properties in both Korean and American adolescent samples.

A History of the MMPI

MMPI. The original MMPI (Hathaway & McKinley, 1943) was developed to provide clinicians with a more efficient and reliable method of diagnostic assessment. Scales were constructed using the empirical keying approach whereby items were determined by differentiating item endorsement based on groups of individuals with a particular presentation of psychopathology versus those without (Graham, 2012). This was in contrast to the popular logical keying approach at the time, which relied on the rationale of the test author's judgment. A benefit of the empirical keying approach was that it prevented difficulties with subjective interpretations of responses and inconsistent item endorsement between groups. After selecting a wide variety of personality-type statements from diverse sources, criterion groups were created from patients, students, and workers affiliated with the University of Minnesota (Graham, 2012). Individuals were divided into groups based on their presenting diagnosis or lack of psychopathology. Subgroups of clinical participants formed the initial MMPI Clinical Scales, which include Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychopathic Deviate (Pd), Paranoia (Pa), Psychasthenia (Pt), Schizophrenia (Sc), and Hypomania (Ma). The Masculinity-Femininity (Mf) and Social Introversion (Si) scales were later added to identify homosexual and

heterosexual men and social personality traits, respectively, completing the construction of the MMPI clinical scales (Drake, 1946).

Due to the notion that test takers could falsify or distort their responses to over or under report their psychopathology, Hathaway and McKinley (1943) also developed the Validity Scales, which included the Cannot Say (?), Infrequent (F), Lie (L), and Correction (K) scales. The Cannot Say scale indicated the number of omitted items or items responded to as both true and false. The F scale identified test takers who responded to items in a deviant manner, and were infrequently endorsed by those in the *normal* subgroup. The L scale detected test takers who attempted to present themselves in an overly favorable way. The K scale (Meehl & Hathaway, 1946) was designed to detect test takers who presented as clinically defensive, by endorsing items that denied psychopathology to appear in a favorable light. The K scale was also used as a form of correcting scores on some of the clinical scales to reflect deviant responding.

Although the MMPI was originally intended as a diagnostic tool for psychopathology, it became apparent that many of its clinical scales were intercorrelated and therefore was not successful in providing valid diagnoses. Over time, the MMPI became an instrument to assess for an individual's overall psychopathology and how one's scale elevations could provide further details on his or her presenting problems. Research also indicated that its originally intended purpose would not have been as useful to gather a range of personality characteristics (Graham, 2012). For example, the results of an individual who elevated only the schizophrenia scale would be less helpful than results that indicated additional comorbid issues. Therefore, while the use of the MMPI changed, it provided a new level of assessment for psychological and personality testing.

While the MMPI became the most frequently used personality test in the United States (Harrison, Kaufman, Hickman, & Kaufman, 1998; Lubin, Larsen, & Matarazzo, 1984), critics expressed concerns about the original sample. Characteristics of the original sample mostly portrayed a 35-year old, white, married worker residing in a small rural town with approximately eight years of formal education, which was not representative of the average American citizen, given the changes in demographics and culture in the United States. There were also concerns with the language and references in the items becoming archaic, sexist, or obsolete (Graham, 2012). Therefore, in 1982, the University of Minnesota Press appointed a restandardization committee to revise the MMPI.

MMPI-2. Major changes encompassed in the MMPI-2 (Butcher et al., 1989) included the rewording of items, the addition of new items, and a new normative sample. Item language was changed to reflect the more contemporary society and new items were added for additional personality and behavioral characteristics including drug abuse, suicide potential, Type A behavior patterns, marital adjustment, work attitudes, and treatment amenability. 2600 participants (1138 men and 1462 women) were selected and tested from seven states across the United States to ensure geographic representativeness and ranged in age from 18 to 85 years old (M = 41.04, SD = 15.29). Ethnically, the sample was composed of Caucasians (81%), African Americans (12%), Hispanics (3%), Native Americans (3%), and Asian Americans (1%). Most men (61.6%) and women (61.2%) were married and had a median income of $30,000-$35,000 for men and $25,000-$30,000 for women. The major goals of the revision were to establish a more representative sample and update item language and content but preserve the meaning and objective of the original MMPI.

Though published in 1989, the MMPI-2 underwent several subsequent developments worth noting. For example, the Infrequency Psychopathology (Fp; Arbisi & Ben-Porath, 1995) and Superlative Self-Presentation (S; Butcher & Han, 1995) scales were added, which detect infrequent responding based on psychiatric inpatient norms and the assessment of individuals who present themselves as highly virtuous and responsible, respectively. In addition, the Content Component scales (Ben-Porath & Sherwood, 1993) were included to aid with the interpretation of the Content scales. Further, the Personality Psychopathology Five (PSY-5; Harkness, McNulty, Ben-Porath, & Graham, 2002) were also included.

MMPI-2-RF. Research indicated that the original clinical scales were heterogeneous in content and strongly intercorrelated (Dahlstrom & Welsh, 1960; Welsh, 1956). Factor analytic studies (Butcher et al., 1989; Eichman, 1961) found that the main source of variance among the Clinical scales of the MMPI and MMPI-2 was the factor of anxiety, general maladjustment, or emotional distress. This emotional distress appeared to be reflected in all of the Clinical scales making it difficult to distinguish and interpret the extent that high elevations accounted for emotional distress or each core construct of the Clinical scales. Tellegen and colleagues (2003) developed the Restructured Clinical (RC) scales, and to aid with more direct interpretations, the Demoralization (RCd) scale. This addition removed the overall emotional distress from the RC scales, which decreased their intercorrelations and increased overall discriminant validity. Along with the development of the nine RC scales, Demoralization (RCd), Somatic Complaints (RC1), Low Positive Emotions (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9), major changes for the revised version of the MMPI-2

Restructured Form (MMPI-2-RF) also included seven revised Validity Scales, one new Validity Scale, two new Interest Scales, and revised versions of the PSY-5.

While the Validity Scales were part of the MMPI-2, the revisions made for the MMPI-2-RF made the scales more independent of each other while still preserving their original intent. The Variable Response Inconsistency-Revised (VRIN-r) scale consists of 53 item-response pairs that detect the number of inconsistent responses to those pairs. The True Response Inconsistency-Revised (TRIN-r) scale consists of 26 pairs of negatively correlated items that calculate the number of inconsistent true or false responding. The Infrequent Somatic Responses (Fs) scale contains 16 items to detect the reporting of uncommon somatic symptoms. The Symptom Validity Scale-Revised (FBS-r) contains 30 items and identifies the likelihood of noncredible reporting of cognitive deficits.

The Higher-Order (H-O) scales were developed to illustrate the overall dimensions of the MMPI-2-RF and include Emotional/Internalizing Dysfunction (EID), Thought Dysfunction (THD), and Behavioral/Externalizing Dysfunction (BXD; Tellegen and Ben-Porath, 2008). Ben-Porath and Tellegan (2008b) developed the Specific Problems (SP) scales to assess for additional clinical concepts not measured by the RC scales. There are 23 scales in total, which assess for somatic (Malaise, Gastrointestinal Complaints, Head Pain Complaints, Neurological Complains, Cognitive Complaints), internalizing (Suicidal/Death Ideation, Helplessness/Hopelessness, Self-Doubt, Inefficacy, Stress/Worry, Anxiety, Anger Proneness, Behavior-Restricting Fears, Multiple Specific Fears), externalizing problems (Juvenile Conduct Problems, Substance Abuse, Aggression, Activation), interpersonal characteristics (Family Problems, Interpersonal Passivity, Social Avoidance, Shyness, Disaffiliativeness), and interests (Aesthetic-Literary, Mechanical-Physical).

MMPI-A. While the MMPI was original intended to assess adults, it was also a popular measure for adolescents. However, there were concerns about the use of the MMPI with adolescents since the norms were standardized for adults. Research also suggested that adolescents who used the MMPI were often overpathologized (Archer, 1984, 1987; Klinge, Lachar, Grisell, & Berman, 1978). Although Marks and Briggs created unofficial adolescent norms in 1967 for MMPI use, the literature indicated inconsistencies in underpathologizing adolescents in the clinical setting (Archer, Stolberg, Gordon, & Goldman, 1986; Dahlstrom, Welsh, & Dahlstrom, 1972; Klinge & Strauss, 1976). Due to these concerns, the MMPI-A was developed specifically for adolescents between the ages of 14 and 18 (Butcher et al., 1992).

After selecting and rewording items from the original MMPI, 478 items were included in the MMPI-A. New items were added that assessed for treatment compliance, attitudes toward changing one's own behavior, treatment-related characteristics, and problems with alcohol and drugs. Items were also added to specifically address adolescent lifestyle such as school behavior, attitudes toward teachers, peer-group influences, eating problems, and relationships with parents and other adults. In total, the MMPI-A consisted of eight validity scales (Cannot Say, VRIN, TRIN, F, F1, F2, L, K), ten Clinical scales (Hypochondriasis, Depression, Hysteria, Psychopathic Deviate, Masculinity-Femininity, Paranoia, Psychastenia, Schizophrenia, Hypomania, Social Introversion), six Supplementary scales (MacAndrew Alcoholism, Alcohol/Drug Problem Proneness, Alcohol/Drug Problem Acknowledgement, Immaturity, Anxiety, Repression), 15 Content scales (Anxiety, Obsessiveness, Depression, Health Concerns, Bizarre Mentation, Anger, Cynicism, Alienation, Conduct Problems, Low Self-Esteem, Low Aspirations, Social Discomfort, Family Problems, Social Problems, Negative Treatment Indicators), and the PSY-5 scales. The F scale was separated into two parts, with F1 providing

validity scores for the first half of the items and F2 providing validity scores for the second half. In addition, Harris-Lingoes subscales and Content Component scales were included to help with the interpretation of the Clinical and Content scales, respectively.

The MMPI-A was standardized from an adolescent normative sample consisting of 805 boys and 815 girls aging from 14 to 18 years randomly selected from seven states across the United States. Although the ethnic demographics were not reported, the MMPI-A manual indicated that the sample was representative of the U.S. population in terms of ethnicity and socioeconomic status. Since there were only minor differences between age groups, there was one set of norms for boys and another for girls. To maintain consistency with the MMPI-2, the raw scores on the eight Clinical scales, the Content and Content Component scales, and the PSY-5 scales were transformed to Uniform T scores.

In comparison with the adolescent norms of the MMPI, the MMPI-A norms indicated significantly lower T scores for most Clinical scales (Mark et al., 1974). Further research found that this difference was greater than five points for adolescent psychiatric patients (Janus, Tolbert, Calestro, & Toepfer, 1996). Butcher et al. (1992) recommended that while T-scores of 65 or greater would be considered clinically significant, T scores that ranged from 60 to 65 could also be interpreted as high scores. Another major concern that arose was that adolescents who presented with clinical problems did not appear to elevate the scales on the MMPI-A. Hilts and Moore (2003) found that 30% of male and 25% of female psychiatric inpatients produced valid MMPI-A results with no elevations in the Clinical scales. One possible explanation for this is the absence of the K-correction that could lead to lower scores for those who were defensive in their testing. However, Alperin, Archer, and Coates (1996) found that even when using the K-correction, there was not a significant difference between test scores. Another possibility is that

there were not many items that differentiated between the normative sample and adolescents in the clinical setting (Archer, Handel, & Lynch, 2001). Since approximately 12% of adolescents in the normative sample reported having been referred to a therapist within six months of completing the MMPI-A, Hand, Archer, Handel, and Forbey (2007) revised the MMPI-A norms to eliminate those individuals. As a result of this research, caution was suggested in MMPI-A interpretation because of possible underpathologizing.

Given that the MMPI-A maintained continuity with the MMPI Validity and Clinical scales in addition to creating new scales for adolescent psychopathology, it gained much popularity and became the only self-report objective assessment of personality in the top 10 instruments for adolescents and was the most frequently used self-report measure in child-related forensic evaluations (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Archer & Newsom, 2000). However, due to maintaining the original MMPI Clinical scales, the criterion-keying method presented its own psychometric limitations. These limitations included inter-related problems of multidimensionality, content heterogeneity, and extensive item overlap between scales, which resulted in excessive intercorrelations and a lack of discriminant validity within the scales. In addition, the length of the MMPI-A continued to pose difficulties for the attention span and concentration for some adolescents (Archer, 2005).

MMPI-A-RF. Given these limitations, the MMPI-A-RF (Archer et al., 2016) was developed to address the problem of heterogeneity through the following approach: (1) develop the demoralization scale to reduce item overlap and intercorrelations, (2) identify the major components of the Clinical scales apart from the demoralization factor using exploratory factor analysis, (3) develop additional adolescent substantive scales, (4) revise and develop Validity Scales for over-reporting, underreporting, and non-content based responding; and (5) revise the

PSY-5 Scales. Using the MMPI-2-RF as a template, the MMPI-A-RF was constructed to maintain continuity between the scales of the two tests. The final developmental sample consisted of 15,128 adolescents (9,286 boys and 5,842 girls) from a variety of settings including inpatient (*n* = 419), outpatient (*n* = 11,699), correctional (*n* = 1,756), and school settings (*n* = 1,254). In terms of ethnicity, the sample was composed of 1,229 Whites (76.3%), 199 Blacks (12.4%), 46 Asians (2.9%), 46 Native Americans (2.9%), 33 Hispanics (2.0%), 41 Other (2.5%), and 16 who did not report ethnicity (1.0%). Ranging in age from 14-18 years, the mean age for the combined samples was 15.61 (*SD*=1.8). Further, the sample was divided into four subgroups, younger boys (14-15), older boys (16-18), younger girls (14-15), and older girls (16-18), to account for developmental factors affecting scale construction.

The MMPI-A-RF consists of 241 items, reduced from the 478-item MMPI-A, and 48 scales overall. It contains six Validity Scales: VRIN-r (random responding), TRIN-r (fixed responding), Combined Response Inconsistency (CRIN; combination of fixed and random responding), F-r (responses infrequent in the general population), Uncommon Virtues (L-r; rarely claimed moral attributes or activities), and Adjustment Validity (K-r; uncommonly high level of psychological adjustment). Three Higher-Order (H-O) Scales were also included, consisting of the Emotional/Internalizing Dysfunction (EID; problems associated with mood and affect), Thought Dysfunction (THD; problems associated with disordered thinking), and Behavioral/Externalizing Dysfunction (BXD; problems associated with undercontrolled behavior). The RC scales were constructed in continuity with the MMPI-2-RF and have identical scale names.

With the goal of maintaining continuity with the MMPI-2-RF, the MMPI-A-RF Specific Problem Scales were developed by first aligning the MMPI-2-RF Specific Problem Scales items

onto corresponding MMPI-A items. A factor analysis was then calculated to analyze the various sets of items including the 58 items unique to the MMPI-A item pool. When a preliminary set of Specific Problems Scales was established, another factor analysis was calculated on the items of each Specific Problem Scale with the items from RCd to ensure that the items that were strongly correlated with RCd were not included in many of the Specific Problem scales, although some items shared the same construct. The Specific Problem Scales were further refined by removing items that were highly correlated with other scales. Finally, the scales were correlated with all 478 MMPI-A items to add items that were relevant in content but not correlated with other scales. All 25 Specific Problems Scales do not overlap in items. Although many of the scales are brief, and thus not surprisingly have relatively low Cronbach's alpha coefficients, the standard errors of measurement (SEMs) were adequate. However, the literature suggests that Cronbach's alpha coefficients under .70 do not always reflect inadequate reliability when SEMs are also evaluated (Schmitt, 1996; American Educational Research Association, 2014). There are 19 of the 25 Specific Problem scales that have direct counterparts to the MMPI-2-RF, exceptions are the Obsessions/Compulsions (OCS), Negative School Attitudes (NSA), Antisocial Attitudes (ASA), Conduct Problems (CNP), Negative Peer Influence (NPI), and Specific Fears (SPF) scales.

The Specific Problem Scales were developed in four sets including Somatic/Cognitive, Internalizing, Externalizing, and Interpersonal scales. The Somatic/Cognitive scales consist of Malaise (MLS; overall sense of physical debilitation), Gastrointestinal Complaints (GIC; nausea, recurring upset stomach, poor appetite), Head Pain Complaints (HPC; head and neck pain), Neurological Complaints (NUC; dizziness, weakness, paralysis, loss of balance), and Cognitive Complaints (COG; memory problems, difficulties concentrating). The Internalizing Scales

include Helplessness/Hopelessness (HLP; belief that goals cannot be reached or problems solved), Self-Doubt (SFD; lack of self-confidence, feelings of uselessness), Inefficacy (NFC; belief that one is indecisive and inefficacious), Obsessions/Compulsions (OCS; varied obsessional and compulsive behaviors), Stress/Worry (STW; preoccupation with disappointments, difficulty with time pressure), Anxiety (AXY; pervasive anxiety, frights, frequent nightmares), Anger Proneness (ANP; easily angered, impatient with others), Behavior-Restricting Fears (BRF; fears that significantly inhibit normal behavior), and Specific Fears (SPF; multiple specific fears). The Interpersonal Scales consist of Family Problems (FML; conflictual family relationships), Interpersonal Passivity (IPP; being unassertive and submissive), Social Avoidance (SAV; avoiding or not enjoying social events), Shyness (SHY; feeling uncomfortable and anxious around others), and Disaffiliativeness (DSF; disliking people and being around them). The PSY-5-r Scales were also included and mimic the MMPI-2-RF.

MMPI-A-RF Externalizing Scales. The MMPI-A-RF contains six Externalizing Scales that measure aspects of two RC Scales: Antisocial Behavior (RC4) and Hypomanic Activation (RC9). The Negative School Attitudes (NSA), Antisocial Attitudes (ASA), Conduct Problems (CNP), Substance Abuse (SUB), and Negative Peer Influences (NPI) scales are facets of RC4, while the Aggression (AGG) scale measures a component of RC9. The Externalizing Scales can also be used to interpret RC3. The Externalizing Scales specify issues presented even when the RC scales may not be elevated. For example, if RC4 is not clinically significant but CNP and SUB are, the Externalizing Scales are still interpretable and clinically significant. Since the current study focused on the Externalizing Scales of the MMPI-A-RF, they are discussed in more detail below.

The Antisocial Behavior (RC4) scale contains 20 items that assess for aspects of disordered and antisocial conduct. An elevated RC4 score (T=60 and above) can be associated with problems at school and home, a tendency to affiliate with socially undesirable peer groups, and issues with alcohol and drugs. Low RC4 scores (T=40 and below) indicated a reduced risk for disordered conduct. The Hypomanic Activation (RC9) scale assesses for aspects of overactivation, aggression, impulsivity, and grandiosity. An elevated RC9 score can be correlated with anger problems, impulsive behaviors, and an inflated ego. Low RC9 scores indicate underactive behaviors, a passive nature, and reduced impulsivity. The Cynicism (RC3) scale consists of nine items describing a negative and cynical view of human nature and interpersonal relationships. An elevated RC3 score can be interpreted to portray an individual who lacks trust in the behavior and intentions of others. Low RC3 scores can be associated with a trustful and well-intentioned view of others.

The Negative School Attitudes (NSA) scale consists of six items describing attitudes and beliefs that school is unproductive and aversive. High scores on this scale can be interpreted to portray one who thinks school is boring, a waste of time, and would choose to avoid school if possible. Low scores can be associated with a favorable attitude of school and its activities. The Antisocial Attitudes (ASA) scale contains six items describing antisocial beliefs and attitudes. Elevated scores can be associated with being entertained by the cleverness of criminals, a tendency to bend or break rules, and being dishonest when met with conflict. Low scores are associated with a below-average number of antisocial attitudes. The Conduct Problems (CNP) scale consists of seven items describing a history of conduct issues in school and at home. Elevated scores are correlated with significant behavioral and academic problems at school and at home, whereas low scores can be interpreted as a history of good behavior.

The Substance Abuse (SUB) scale consists of four items describing drug and alcohol use problems. Elevated scores can be associated with an increased risk for substance abuse. Items in this scale have been identified as having critical item content to be brought to clinicians' immediate attention, and are marked in the scoring report if endorsed in the keyed direction. The Negative Peer Influence (NPI) scale is comprised of five items describing an association with peers who encourage and support antisocial behaviors. High scores indicate issues related to affiliation with socially undesirable peer groups. The Aggression (AGG) scale contains eight items describing aggressive behaviors and attitudes. Elevated scores indicate engaging in physically aggressive, violent behavior, and the enjoyment of intimidating others. Low scores can be associated with below-average aggression. Items in this scale have also been marked as critical item content.

Korean Translations of the MMPI

The family of MMPI assessments has been translated into many different languages. The MMPI-2 has been translated into 21 different languages, and the MMPI-2-RF was translated into four languages. In addition, the MMPI-A has been translated into 13 languages. The MMPI, MMPI-2, MMPI-A, and MMPI-2-RF all have been translated into the Korean language (Cheung, Lee, & Jin, 1963; Han, 1993; Han, Moon, Lee, & Kim, 2011; Kim et al., 1989). To illustrate the validation process of translating an MMPI assessment into the Korean language, the translation process of the MMPI-2 will be used as an example.

The MMPI-2 was first independently translated into Korean by Han (1990) and another bilingual student. After comparing their two versions for further discrepancies, the Korean items were given to another bilingual student in order to back translate them into English. An American MMPI psychologist who was an expert in MMPI cross-cultural studies then examined

the original English version and the back translation. The psychologist found that 20 items were non-equivalent between the original English and back translated version, subsequently, the items were reviewed, re-translated, and judged to be equivalent to the original English version (Han, 1993).

The Korean MMPI-2 was then administered to 726 Korean college students from eight universities across Korea to examine its psychometric properties. The validation process consisted of analysis of its factor structure, 1-week test-retest, and peer behavior correlates (Han, 1993). In final efforts to review its validity, 53 bilinguals living the United States completed the English and Korean versions of the MMPI-2 within a 1-week interval to examine test-retest reliability. The results supported the equivalence of the Korean MMPI-2 with similar mean profiles, scale score correlations, and comparable magnitude in test-retest correlations (Chung, Weed, & Han, 2006).

Juvenile Delinquency in Korea and the United States

The underlying construct of antisocial personality disorder consists of three major components: criminal behaviors, interpersonal antisocial behaviors, and intrapersonal antisocial behaviors (Dinges, Atlis, & Vincent, 1997). Criminal behaviors can be described as behaviors that are deviant from legal norms. Interpersonal antisocial behaviors can be defined as behaviors that create conflict with other individuals in a social environment, while intrapersonal antisocial behaviors can occur within an individual, for example, impulsivity or a lack of self-regulation. These aspects of antisocial personality disorder, or conduct disorder in adolescents, are captured by the externalizing scales of the MMPI-A-RF. However, the broad components of antisocial behaviors may have different underlying facets, which can be weighted with varying emphasis across cultures (Dinges et al., 1997).

An overarching concept that distinguishes between Eastern and Western cultures is the role of individualism and collectivism. Individualism can be portrayed in a society where individuals conceive themselves to be autonomous and distinct from others, orienting their values and behaviors towards oneself and internally. Collectivism can be illustrated through the perception of the self as embedded within social roles and relationships, orienting values and behaviors towards one's group membership or family (Le & Stockdale, 2005). Western cultures tend to be more individualistic, whereas Asian cultures tend to be more collectivistic (Singelis, Triandis, Bhawuk, & Gelfand, 1995). The individualistic and collectivistic perspectives heavily affect the interpretation and weight of values and behaviors embedded across cultures, especially externalizing behaviors since they commonly involve and impact those surrounding one's self.

Negative School Attitudes (NSA) Interpretation. In the Korean culture, the philosophy of Confucianism greatly influences all aspects of society including the home, school, community, and political context (Oh-Hwang, 1993). Confucianism places filial piety that is, respect for one's parents, elders, and ancestors, as the foundation of all conduct, which affects all other interpersonal relationships. Koreans are taught proper interpersonal relations, which are grounded in the five basic hierarchical human relationships in Confucianism: (1) ruler and subject; (2) father and son; (3) husband and wife; (4) older brother and younger brother; and (5) between friends (Reagan, 1996). While Western society continues to make a growing impact on Eastern values, Korea is still known to be a patriarchal society. Thus, this hierarchy of relationships can be seen across many aspects of Korean society including its educational structure.

The Korean culture places the value of education as an important element in society due to the principles of Confucianism. Considering the hierarchy of relationships in Confucianism,

the role of an educator can be identified as a ruler at school and is therefore revered as an authority figure in Korea (Shin & Koh, 2005). Students are motivated to uphold high educational standards because education raises their socioeconomic status and upholds their family's honor (Siu, 1992). From a collectivistic standpoint, high educational attainment not only benefits the individual but also compliments the family as a whole. However, with only a limited number of college placements and job opportunities, the educational culture is extremely competitive, creating a large amount of stress for students (Shin & Koh, 2005).

Therefore, in the context of Korean culture and education, the NSA scale in the MMPI-A-RF may be interpreted as not only having negative attitudes and beliefs about school but may also be associated with the hierarchal structure of education and the adolescents' stress in providing for their families. In the American culture, while the family structure is emphasized, each individual is most benefited or impeded by his or her educational achievement, rather than the collective whole. Koreans with negative attitudes and beliefs about school may be discouraged and stressed by their collectivistic pressures and limited opportunity for success.

Antisocial Attitudes Scale (ASA) and Conduct Problems Scale (CNP) Interpretation. Although the delinquency rates among Korean youth declined in the years up to 1998, between 2004 and 2009, they have dramatically increased, almost doubling in number. According to the Supreme Prosecutors' Office of Korea (2011), in 2010, 87,766 Korean youth were arrested for theft/fraud (45.1%) and assault/violence (25.9%), approximately 30% higher than in 2006. The rates of severe crimes such as murder, rape, and robbery continue to remain low (3.5%), but have been increasing gradually since 2007. Interestingly, according to the National Center for Juvenile Justice (2015), the delinquency rates among adolescents in the United States have reached a historical low point. From 1996 to 2014, the number of juvenile arrests declined by 65%. While

Eastern and Western views of both antisocial attitudes and conduct problems appear to have similar characteristics, describing the risk and protective factors of juvenile delinquency may give further insight to any potential differences.

Moffitt (1993) described that the development of antisocial behaviors in adolescence revolved around the maturity gap. The maturity gap refers to the phenomenon in which adolescents desire independence from their adult guardians, but cannot achieve it due to their social environment. Most delinquency periods are short in duration, with only a small number of juvenile offenders carrying their delinquent behaviors and crimes into adulthood. In a meta-analysis, Cottle, Lee, and Heilbrun (2001) found that age at first offense, family issues, peer groups, and family history of criminality were the strongest predictors of juvenile recidivism in the United States. However, other research (Onifade et al., 2011) found that the notion of delinquency was a cross-product of both proximal and distal factors. While the trends of delinquency share common themes across cultures, Kim and Kim (2008) indicate that there are sociocultural factors related to intergenerational conflict as Confucian values become less prominent in younger generations. Further, although multi-faceted, they believe that the rise in delinquency in Korea can be related to specific cultural factors.

The literature (Blum & Libbey, 2004; Lee & Jun, 2009; Lee, Onifade, Teasley, & Noel, 2012) indicated that parental monitoring, parent attachment, teacher attachment, and academic achievement were important protective factors against juvenile delinquency while delinquent peers and poor parenting were risk factors. Parental monitoring can be described as the parent's knowledge of the social activities of their children as well as their children's friends. It can also include the parent's rules and regulations regarding their children's behavior (Hirschi, 1969). Parent attachment refers to the emotional connection between a parent and their children. When

this emotional connection is low, the expectations parents have for their children's behaviors are also low (Dorius, Bahr, Hoffmann, & Harmon, 2004). While parental supervision and positive parent-child relationships appear to be protective factors against juvenile delinquency in both Korean and American youth (Furstenberg et al., 1999; Gorman-Smith et al., 2000), teacher attachment and academic achievement may have greater emphasis in the Korean culture due to the impact of education on Korean society (Lee & Jun, 2009). Thus, while the ASA and CNP scales appear to have similar implications in both Korean and American youth, analyses and interpretations should account for the cultural expectations of collectivistic values and how these impact antisocial and conduct disorder behaviors.

Substance Abuse (SUB) Interpretation. For the SUB scale on the MMPI-A-RF, interpretations should be made from a cultural perspective due to the restrictions and cultural context of substance use in Korea. The Korean government prohibited the usage and possession of drugs, including marijuana, in 1946, and as a result, substances are available but very difficult to access in Korea compared to their accessibility in the United States (Kwon-Ahn, 2001). However, alcohol consumption in Korean adolescents has increased since 1990 and remains a serious issue in Korean schools (Kim & Kim, 2008). The prevalence rates of alcohol consumption by Korean high school students (44.5%) are similar to the prevalence rates in American high school students (48.6%; Johnston, O'Malley, & Bachman, 2003; Kim & Kim, 2002). Although legal minors (individuals under 20 years) are not able to purchase alcohol legally, it is very easy for minors to purchase alcohol from numerous small retail outlets that are more lenient (Lee, 1997). Since substances apart from alcohol are difficult to access in Korea, Korean youth may abuse alcohol as a replacement.

Comparing between samples of Korean and American adults on the MMPI-2-RF, Wang (2014) found that Koreans yielded a lower overall endorsement of substance abuse items, including alcohol, than Americans. Wang (2014) indicated that the lower level of alcohol endorsement might be explained by the great tolerance of heavy drinking in the Korean culture as Koreans are encouraged to drink alcohol socially and drinking is often viewed as a bonding experience in interpersonal relationships (Kwon-Ahn, 2001). Given the restrictions and cultural context of substances in Korea, interpretations of the SUB scale should take into account these differences.

Negative Peer Influence (NPI) Interpretation. Social learning theory suggests that youth are highly influenced by the behaviors of their peers and other important figures in their lives (Lee et al., 2012). Given the emphasis on the hierarchal structure of relationships in Confucianism, social learning theory plays an important role in the Korean culture by the positive or negative influences of parental relationships, teacher attachment, and peer groups. In the United States, researchers have found that negative peer influence may be the strongest predictor for juvenile delinquency since the majority of crimes committed by adolescents occur in groups (Henry et al., 2001).

In a sample of adolescents living in an inner-city neighborhood in the United States, Henry and colleagues (2001) found that adolescents who experienced low emotional support and inconsistent discipline from their parents, compared to positive emotional support and discipline, reported having more deviant friend groups two years later and more involvement in violent and nonviolent delinquent behaviors five years later. Similarly, Lee and colleagues (2012) found that in Korean youth, an association with delinquent peers resulted in greater levels of delinquent

engagement. Thus, the NPI scale appears to have consistent interpretations in both Korean and American youth.

Aggression (AGG) Interpretation. Aggression can be defined as a behavioral and emotional response that results in destroying or damaging other persons or things (Kim & Kim, 2007). Factors in Korea that can contribute to juvenile aggression include depression, academic stress, and low self-esteem. Due to strong desires for recognition, friendship, affection, independence, dependence, success, pride, and self-identity in adolescents, frustration and depression can arise if these needs are not met (Kim & Chung, 2004). Adolescents may behave aggressively as a negative coping strategy to serve as an outlet for their frustration and feelings of depression (Kim & Lee, 2008). In addition, since all Korean students are impacted by the strongly emphasized education system, academic stress can serve as the starting point of every juvenile issue in Korea, which can ultimately lead to aggression as a coping mechanism (Park, Choi, & Lim, 2014). Another factor influencing aggression is low self-esteem. Studies (Wang et al., 2013; Webster & Kirkpatrick, 2006) have found that adolescents with high self-esteem show better interpersonal relationships and decreased aggressive behaviors, while adolescents with low self-esteem tend to have higher levels of aggression. In a sample of Korean middle school students, Park and colleagues (2014) found that academic stress, depression, self-esteem, decision-making competency, and happiness were correlated with aggression.

The literature in the United States also indicates similar factors influencing the onset of aggression in juvenile delinquents. Both social learning theory and environmental demands foster the development of individual beliefs and behaviors like aggression given contextual circumstances (Hawley, 2003). Ng-Mak and colleagues (2004) found that exposure to stressful events, such as community violence, has been associated with beliefs supporting violence and

aggression. Given the influence of parental attachment and interpersonal relationships, Orpinas and colleagues (1999) found that youth's perception of parental support for fighting was also associated with aggressive behaviors. In addition, Rappaport and Thomas (2004) found that deviant peer affiliations were a positive predictor of aggression in American youth. While the stressful events in Korea and America may be different, the factors that lead to aggression appear to be consistent. Therefore, the AGG scale on the MMPI-A-RF should account for the similar construct of aggression in Korean and American adolescents.

Korean and American MMPI Comparisons

While this study focuses on the MMPI-A-RF, because of the limited number of studies between the Korean and American populations, the following literature review will focus on the cross-cultural research of the MMPI, MMPI-2, MMPI-2-RF, and MMPI-A to illustrate an overview of the literature in Korean and American cultures.

Using the Korean and American normative samples of the MMPI-2 and MMPI-A, Han and colleagues (2014) examined whether there were gender differences in the items and content domains across cultures. They hypothesized that Americans would show a greater gender difference compared to Koreans due to the gender roles of the MMPI assessments being standardized in America, with adults indicating a greater gender difference than adolescents by more traditional gender roles having been established through aging. By calculating the percentages of responding "true" to each MMPI item separated by gender, and obtaining the endorsement percentage difference between genders, they found, as expected, that American adults (42 items; 7.4% of the item pool) had a significantly greater proportion of gender-discriminating items compared to Korean adults (17 items; 3.0%). Although American adolescents (21 items; 4.4%) had a greater proportion of gender-discriminating items compared

to Korean adolescents (13 items; 2.7%), the difference was not significant. When differentiating between content domains across cultures, both groups showed very similar content dimensions revolving around stereotypical gender interests, behaviors, and emotions.

By analyzing the normative samples of the MMPI-2 in Korea and America, Ketterer (2010) used multiple-group CFA to examine the measurement invariance of the Restructured Clinical Scales on the MMPI-2 across cultures. The author found that the one-factor model of RC1, RC2, RC3, RC4, RC7, and RC9 demonstrated poor overall model fit across Korean and American samples. However, the author found that configural invariance was established for RCd, RC6, and RC8 scales indicating that Demoralization, Ideas of Persecution, and Aberrant Experiences were constructs that were similarly defined across groups. In addition, partial factorial invariance was established for RCd and RC6 and fully supported for RC8 indicating that the Demoralization and Ideas of Persecution had similar meaning and structure while Aberrant Experiences had identical meaning and structure across cultures. Ketterer (2010) concluded that future research on the factor structures was warranted.

Wang (2014) examined the measurement invariance of the MMPI-2 Restructured Clinical Scale RC4 across Korean and American clinical samples using multiple-group CFA. Expanding on Ketterer's (2010) study of the measurement invariance of the Restructured Clinical Scales on the MMPI-2, the author used the four-factor model (School Problems, Substance Abuse, Family Problems, Violation of Social Norms) of RC4 rationally derived from Han and colleagues (2011), as opposed to the one-factor model developed in the MMPI-2. The author found that the four-factor model showed better model fit overall compared to the one-factor model across cultures. The results indicated partial scalar invariance, in that only the Family Problems factor showed full scalar invariance between groups. In addition, three latent factors (School Problems,

Substance Abuse, Violation of Social Norms) yielded a lower endorsement of items in the Korean sample compared to the American sample. More specifically, on the School Problems factor, Koreans yielded a lower endorsement of item 223 (school suspension). For the Substance Abuse factor Koreans yielded a lower endorsement in the keyed direction of items 237 (never used prescription drugs), 49 (like using marijuana), and 297 (get drunk frequently). Again, in accordance with cultural values, on the Violation of Social Norms factor, Koreans yielded a lower endorsement on items 38 (never having problems because of sex behavior), 190 (never having problems with the law), and 21 (stole things at a young age). The author concluded that while RC4 was appropriate for assessment of Antisocial Behavior in the Korean clinical settings, researchers should be cautious of the items that yielded lower endorsement in the Korean sample.

Han and Lim (2001) used the MMPI-2 to compare 167 Korean college students and 120 Korean psychiatric patients with the American normative and clinical samples, respectively. They found that Korean college students showed significantly elevated mean scores compared to the American normative sample. Within the Korean samples, psychiatric patients produced moderately higher mean elevations than college students as compared to the significant difference in mean scores found between the American normative and clinical sample. However, when comparing Korean and American clinical samples, they found only a small difference between mean scores. The authors concluded that the range of the Korean psychiatric sample scores might have been more restricted due to a ceiling effect. In addition, they inferred that the MMPI-2 items do not discriminate as well between normal and clinical samples in Korea compared to the discrepancy in American samples.

Wang (2017) used multiple-group CFA to examine the measurement invariance of the MMPI-2-RF Externalizing Specific Problems Scales (Juvenile Conduct Problems, JCP; Substance Abuse, SUB; Aggression, AGG; Activation, ACT) using American and Korean normative and clinical samples. First, the author tested the measurement invariance of these scales across genders for each culture and found that all Externalizing Specific Problem scales exhibited partial scalar invariance with some gender noninvariant items for all American clinical and normative samples while for Korean samples, most of the scales showed full measurement invariance except for scales ACT and AGG, which only reached partial scalar invariance in the Korean normative sample. Second, measurement invariance was examined across cultures by incorporating gender noninvariant items and found that most of the scales reached partial scalar invariance except for the JCP scale, which reached full scalar invariance for clinical samples.

Establishing Measurement Invariance

Measurement invariance is a notion that an item relates to an underlying construct or latent variable comparably across groups (Millsap, 2011). As an important aspect of test development, it is intended that the measurement properties of tests assessed in a heterogeneous population should be equivalent in subgroups of a population. For example, tests assessing for IQ in males and females should represent the same level of cognitive ability, in that for a given level of true intelligence, males and females should score equally. Questions of measurement invariance can be addressed in confirmatory factor analysis (CFA) by multiple-groups. In a multiple-group CFA, the measurement model is simultaneously estimated in two or more subgroups. Since multiple-group CFA is a specific procedure of CFA, an outline of CFA is discussed below.

Confirmatory Factor Analysis. CFA is a type of structural equation modeling (SEM) that specifically assesses the relationships between observed measures or indictors (e.g., test items, test scores) and latent variables or factors (Brown, 2006). Although similar to exploratory factor analysis (EFA), CFA is hypothesis-driven in that the researcher must have *a priori* sense of all aspects of the CFA model based on strong empirical evidence or theory. Typically, CFA is used after the underlying structure of the measurement model has been established by EFA. Both EFA and CFA use the same estimation method, maximum likelihood (ML), which evaluates how well the factor models reproduce the input variance/covariance matrix, also known as goodness-of-fit.

Traditionally, the variables in an EFA are completely standardized. Specifically, for completely standardized results, factor variances are set to 1.0, factor loadings are correlations or standardized regression coefficients, and both latent factors and indicators are completely standardized. Unlike EFA, the results of CFA can be unstandardized, standardized, and completely standardized. While EFA uses a correlation matrix (i.e., completely standardized variance-covariance matrix), CFA typically uses an unstandardized variance-covariance matrix to produce CFA results (Brown, 2006). An unstandardized CFA solution expresses its results using the original metrics of the indicators. Standardized CFA solutions consist of unstandardized indicators and standardized latent variables. Although CFA results can be reported in completely standardized formats, researchers find that reporting standardized results can be potentially misleading and unstandardized results provide clearer interpretations since the major components of a CFA are based on unstandardized variables (Willett, Singer, & Martin, 1998).

To produce the best possible analyses, CFA additionally allows its parameters to be freely estimated, fixed, or constrained (Brown, 2006). This allows researchers to evaluate the measurement model by comparing whether the fit of a more restricted solution is better with or without constraints. For example, a parent model consists of five indicators that freely load onto one factor. As a subset of the parent model, a nested model consists of the same five indicators but the factor loadings are constrained to load equally onto one factor. An analysis of both models allows researchers to statistically compare and evaluate the conditions of the factor loadings. Freely estimated parameters refer to an analysis that finds the values for the parameters that reproduce the variance-covariance matrix. In contrast, researchers who use fixed parameters assign specific values to the parameters of the measurement model. Lastly, constrained parameters do not assign specific values to the parameters, but identify additional restrictions on the conditions of the parameters such as allowing the factor loadings to equally load onto one factor. Constrained parameters are fundamental in multiple-group CFA since equality constraints are placed on the parameters of two or more groups, allowing researchers to compare model fit with specific factors, which will be discussed in more detail below.

CFA Model Parameters. Factor loadings (Λ), factor variances-covariances (Φ), and unique variances (Θ_δ) are estimated for all CFA models (Brown, 2006). All parameters are reported in unstandardized and standardized forms. Unstandardized factor loadings are the regression slopes for the prediction of indicators from latent factors. Standardized factor loadings are estimated correlations when items are congeneric when modeled with a single factor, or partial regression coefficients when modeled with more than one factor. Unstandardized unique variances, also known as error variances, refer to the raw variance in an indicator unexplained by latent factors and when standardized, represent the relative variance in the indicators not

explained by the latent factors. In an unstandardized solution, factor variances are the overall dispersion of a latent factor, or the variability of the sample on the latent construct and are equal to 1.00 when standardized. Error covariance (δ_{COV}) and factor covariances (Φ_{COV}) may also be included in CFAs when justified. Unstandardized error covariance, the relationship between measurement errors, refers to the partial covariance between two indicators that is not explained by the latent factor and reflects partial correlations when standardized. Finally, unstandardized factor covariances are the estimated relationship between latent factors when two or more factors are used and correlations between factors when standardized.

The CFA parameters previously discussed are calculated to reproduce the input variance-covariance matrix. While the indicators of the analysis of covariance traditionally are measured as deviations from their means, indicator means being equal to zero, the CFA model can also include the analysis of mean structures. In this case, the CFA parameters are expanded to reproduce the observed sample means of the indicators within the input variance-covariance matrix. For example, in multiple-group CFA, the estimates of the indicator intercepts (predicted value of the indicator when the factor is zero) and the latent factor means are calculated to distinguish the relationship between groups on the underlying latent construct.

Goodness-of-fit Indices. The goodness-of-fit indices represent the statistical comparisons of how well measurement models fit or reproduce input data (Brown, 2006). The goodness-of-fit indices that are most commonly used include chi-squared (χ^2), the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI; Brown, 2006). Based on hypothesis significance testing, a statistically significant χ^2 rejects the null hypothesis and supports the alternate hypothesis that the model estimates do not sufficiently reproduce the

sample variance-covariance matrix. Thus, to ensure a valid reproduction of the sample variance-covariance matrix, researchers strive for a non-significant χ^2 difference test. Although the χ^2 difference test is known to be the classic goodness-of-fit index, criticisms of χ^2 are noted due to its absolute fit criteria. First, the underlying distribution of χ^2 is comprised of small sample size data sets. Second, in large size data sets, the χ^2 difference test is inflated, thus commonly mistakenly rejecting the null hypothesis. Third, since it is based on hypothesis significance testing, there is no room for more reasonable and less stringent comparisons, which can also mistakenly reject the null hypothesis.

Another goodness-of-fit index that falls under the absolute fit category is the SRMR. Deriving from a residual correlation matrix and as the positive square root average, the SRMR can be interpreted as the average difference between the correlations in the observed input matrix and the predicted model. It is calculated by summing the squared coefficients of the residual correlation matrix, dividing this sum by the number of coefficients, and taking the square root of this result. Ranging in value between 0.0 and 1.0, the SRMR indicates a perfect fit with a value of 0.0, with smaller values corresponding to better model fit.

The RMSEA is a population-based index that includes a penalty function for poor model fit that favors a fit with fewer freely estimated parameters. Based on the noncentral χ^2 distribution (the distribution of the fitting function for a non-perfect model), it incorporates a noncentrality parameter (NCP) that reports the degree of model misspecification. As opposed to the more stringent χ^2 difference test, RMSEA is an error of approximation index since it expresses how reasonably well a model fits in the population. Similar to the SRMR, the RMSEA's upper range rarely exceeds 1.0 with values of 0.0 indicating a perfect fit and values closer to 0.0 indicating better fit.

Known to be among the best behaved indices in the literature, comparative fit indices including CFI and TLI, assess the fit of the researcher's CFA solution with a more restricted baseline model, or null model. In this baseline model, the covariances within the indicators are equal to 0.0 and no parameter constraints are placed on the indicator variances. Due to the lack of constraints, comparative fit indices are less stringent than the indices previously discussed. The values of CFI range from 0.0 to 1.0, with values closer to 1.0 indicating better model fit. Similar to RMSEA, CFI is also based on the NCP within the noncentral χ^2 distribution. Like the RMSEA, the TLI also associates a penalty function for the addition of freely estimated parameters. Unlike the CFI, the TLI is non-normed, meaning that its range of values exceeds 1.0. However, values closer to 1.0 still indicate a good model fit.

The guidelines for interpreting goodness-of-fit indices are often debated due to the complexities in the diversity of analytic situations including sample size, model variations, estimation method, misspecification, and normality and type of data. However, from their comprehensive simulation studies, Hu and Bentler (1999) provided the following guidelines to interpret goodness-of-fit indices. Reasonably good fit between the target model and observed data is obtained when (1) SRMR values are close to .08 or below; (2) RMSEA values are close to .06 or below; and (3) CFI and TLI values are close to .95 or above. Browne and Cudeck (1993) provided guidelines that follow a range of fit index values rather than specific cutoff points. They suggested that RMSEA values less than .08 indicate adequate model fit and RMSEA values below .05 show good model fit, while RMSEA values equal to or greater than 0.1 should be rejected. In addition, other researchers have suggested that CFI and TLI values lower than .90 should be rejected, while values ranging from .90 to .95 may be acceptable (Bentler, 1990). While these guidelines provide a comprehensive evaluation for goodness-of-fit

indices, researchers should take into account the variations found in different analytic situations. Due to the issues of Type I and Type II error, it is recommended that researchers use a combination of indices to evaluate the fit of CFA models.

Modification Indices. As another component to model evaluation, the modification index is an approximation of the difference in overall χ^2 between a model with fixed or constrained parameters and a model where the parameters are freely estimated (Brown, 2006). Similar to χ^2 difference testing, the modification index evaluates the amount the χ^2 would decrease if the fixed or constrained parameters were freely estimated. The amount the χ^2 decreases may be smaller or larger than the actual modification index. Modification indices can be calculated for each fixed and constrained parameter in the model. Since modification indices can be interpreted as a χ^2 statistic with 1 *df*, a model showing good fit should indicate values that are lower than the critical value of 3.84, (χ^2 at $p < .05$, 1 *df*). Thus, if the modification index is greater than 3.84, this suggests that the model fit with freely estimated parameters would be a significant improvement compared to fixed or constrained parameters.

Since modification indices are influenced by sample size, adding a parameter may be suggested when *N* is very large. Due to an additional parameter creating issues in the model, an unstandardized, standardized, or completely standardized expected parameter change (EPC) value can be calculated to indicate the change in parameter estimates in a positive or negative direction and help interpret whether the respecification is valid. However, freeing a parameter to improve model fit should be validated and supported by prior research or theory.

Multiple-Group Confirmatory Factor Analysis

As previously noted, measurement invariance is the evaluation of across-group equivalence of parameters including factor loadings, intercepts, and residual variances. Multiple-

group CFA refers to the simultaneous analysis of CFA in more than one group (Brown, 2006). When involving two groups, two separate input matrices are assessed while constrains are placed on the parameters to evaluate the equivalence of the measurement model and the structural solution, or population heterogeneity. Population heterogeneity involves the assessment of the structural parameters, which include factor variances, covariances, and latent means. An advantage of multiple-group CFA is that it examines all potential aspects of invariance across groups, which include configural invariance, metric invariance, scalar invariance, and factorial invariance. Configural invariance, or equal form, is the test of equal factor structures to assess whether the number of factors and pattern of indictor-factor structures are uniform across groups. Metric invariance, or equal factor loadings, is the test of the equality of factor loadings across groups. Scalar invariance, or equal intercepts, assesses the equality of indicator intercepts across groups. Finally, factorial variance, or equal residuals, examines the equality of indicator residuals across groups.

While there has been much research on the procedure of multiple-group CFA, the literature recommends a "step-up" approach to invariance evaluation due to the viability of assessing the least restricted solutions (equal form) to determine whether the next steps of multiple-group CFA are viable and warranted (Brown, 2006). The recommended order of multiple-group CFA invariance evaluation is as follows: (1) examine simultaneous tests of equal form; (2) test the equality of factor loadings; (3) test of equality of indicator intercepts; (4) test the equality of indicator residual variances; (5) test of equality of factor variances; (6) test of equality of factor covariances; and (7) test the equality of latent measures. Tests 1 – 4 are the steps to evaluate measurement invariance and tests 5 – 7 determine structural invariance. However, before conducting these series of tests, the CFA model should first be tested separately

in each group to establish that a one-factor model is acceptable in both groups. A difference in measurement models would conflict with further invariance evaluation. Once both groups show good model fit, the tests for measurement invariance can begin.

Multiple-Group Confirmatory Factor Analysis with Categorical Variables

While the sequence of tests discussed above portrays an overview of multiple-group CFA, the cited sequence is only appropriate for continuous indicators. For dichotomous indicators, the variable type in this study, the procedure follows a slightly different format. In the case of treating categorical indicators as continuous indicators, the analysis can attenuate the correlation estimates, and thereby create "pseudofactors," resulting in skewed test statistics and standard errors. For dichotomous or categorical indicators, rather than using a sample variance-covariance matrix, the analysis will be based on a tetrachoric correlation matrix, alternating the procedure of tests (Brown, 2006).

Based on the Mplus framework, latent continuous response variables (y*) are used to arrange for response models for categorical variables (Muthen & Asparouhov, 2002). Within this unifying framework, y* can be interpreted as the amount of an underlying continuous characteristic that is needed to produce a certain response for a categorical variable. For example, in a test for anxiety, y* would reflect the level of the underlying anxiety required to respond in a certain direction on dichotomous test items. While continuous indicators would allow for an interpretation with more precision in the individual differences for anxiety, y* estimates the amount of the underlying construct for a certain direction. In this way, rather than using the correlations of its observed variables, the correlations of the underlying y* variables are analyzed (Brown, 2006).

The underling y* variables are able to estimate observed categorical variables by using threshold parameters. Similar to dummy coding, in terms of dichotomous variables (y = 0 or 1), the threshold is exceeded when y = 1 and the threshold is not exceeded when y = 0. In the case of multiple-group CFA, the threshold is based on the mean structure of the CFA model. In addition, since multiple-group CFA analysis will be based on a tetrachoric correlation matrix for the interpretation of y*s, the observed variances of the indicators are not calculated. Thus, since residual variances of categorical indicators are not analyzed, the measurement errors of the CFA model can be calculated by 1 minus the squared completely standardized factor loading (Brown, 2006). The test sequence for multiple-group CFA with categorical indicators is as follows:

Configural Invariance. First, the test for configural invariance establishes whether identical factor structures exist across groups. Equal factor structures refer to the same number of factors being associated with the same latent factor across groups. Since no invariance constraints are placed in the equal form tests, essentially, the program software "stacks" the two CFA analyses on top of each other. If equal form is established, this solution will serve as the baseline model for the remaining series of tests. Although multiple-group CFA can be calculated with varying group sizes, a large difference in sample sizes can create difficulties within the solution due to its reliance on sample size. Difficulties may arise in the χ^2, modification indices, standard errors, power, and standardized residuals. Researchers must be mindful of the influence of sample size when interpreting results (Brown, 2006).

Factorial Invariance. Second, for multiple-group CFA with categorical indicators, factorial invariance combines metric and scalar invariance testing. Factorial invariance, or the equality and constraining of factor loadings and thresholds, assesses whether the measures have the same meaning and structure across different groups and that the thresholds have identical

origin points (zero point) across groups (Ketterer, 2010). The equality of factor loadings is a crucial step in measurement invariance since the results provide the researcher with the validity of further group comparisons. By constraining the unstandardized factor loadings, if there is not significant decrease in overall model fit, this will conclude that the measures have the same meaning and structure across groups, thus establishing the equality of factor loadings (Brown, 2006). The equality of thresholds can be interpreted as any point on the latent factor, the expected score of an indicator will be equivalent across groups. However, if the thresholds are noninvariant, the mean differences on the latent factor may not be affiliated with mean group differences due to a number of analytic solutions such as flawed measurement or method bias (Brown, 2006; Chen, 2008).

Residual Invariance. Residual invariance is established by placing equality constraints on the residual variances and covariances of each group. This test determines whether the indicators measure the latent factor with the same degree of measurement error. If residual variances and covariances are equivalent across groups, the amount of variance that is unexplained by the latent factor is the same (Cheung & Rensvold, 2002). In other words, if the measurement error across groups is the same, the unaccounted variance from the latent factors comes from similar places in each group (Brown, 2006).

Latent Mean Invariance. Equal latent means are established by placing additional constraints on the factor means. Differences are then observed to examine the average latent factor score between groups. Since multiple-group CFA accounts for the measurement error among its parameters, mean comparisons in multiple-group CFA are more accurate and have more power than traditional significance tests (e.g., *t*-tests, ANOVA; Brown, 2006). If the latent means are invariant across groups, this can be interpreted as the groups have equal average

means on the latent construct. However, if found noninvariant, group differences can provide meaningful interpretations between groups.

CHAPTER II

RATIONALE OF THE PRESENT STUDY

The purpose of the present study is to investigate the measurement invariance of the MMPI-A-RF Externalizing Scales across Korean and American adolescent normative samples. Given the literature on the cultural differences between the Korean and American population, differences in cultural and environmental values may give researchers more insight on juvenile delinquency across cultures. In order to establish whether differences in item endorsement reflect true cultural differences or measurement bias, further testing is warranted to investigate the measurement invariance of the MMPI-A-RF Externalizing Scales. If full measurement invariance (configural, factorial, residual, latent) is found using the multiple-group CFA approach, the differences in item endorsements will reveal the cultural differences between Korean and American adolescent samples. If noninvariance or partial invariance is found, measurement bias in the MMPI-A-RF may be exhibited across cultures.

The study is an important contribution to the literature of this newly developed measure. First, this study investigated whether the underlying constructs and facets of the MMPI-A-RF Externalizing Scales are reflective of each population and carry valid interpretations across cultures. Second, this study provided additional psychometric research to the MMPI-A-RF in both Korean and American adolescent normative samples, furthering the establishment of measurement invariance testing.

In cross-cultural studies using measurement invariance testing, Chen (2008) found that approximately 94% of studies reached the configural invariance stage, meaning that the same number of factors were associated with the same latent factor. As finding full measurement invariance for assessments between cultures can be quite challenging, the results of previous

Korean to American MMPI measurement invariance studies were examined to propose this study's hypothesis. Previously discussed, Ketterer (2010) found that two out of the nine RC scales of the MMPI-2 showed partial factorial invariance while one scale indicated full factorial invariance. When examining the RC4 scale of the MMPI-2, Wang (2014) found that three of its four latent constructs showed partial scalar invariance, while one of the latent constructs showed full scalar invariance. Lastly, Wang (2017) examined the SP scales in the MMPI-2-RF, finding that all of the scales met partial scalar invariance except for the JCP scale, which reached full scalar invariance for clinical samples. Given the results of these studies and Chen's (2008) measurement invariance analysis, this study hypothesizes that all MMPI-A-RF Externalizing Scales would reach partial factorial invariance.

CHAPTER III

METHODOLOGY

Participants

American adolescent normative sample. The American MMPI-A-RF normative sample is a subset of the normative sample for the MMPI-A. It was gathered from middle and high school students throughout several states (i.e., California, Minnesota, New York, North Carolina, Ohio, Pennsylvania, Virginia, and Washington State) to obtain a balanced sample based on ethnicity, geography, and rural-urban residence. Collected in the late 1980s and early 1990s, the MMPI-A normative sample was found to be consistent with contemporary adolescent samples by reviewing studies conducted between 1995 to 2012 that reported MMPI-A Clinical Scale means and standard deviations for nonclinical samples (Archer, in press). These studies indicated that T scores based on the MMPI-A normative sample were appropriate for current MMPI-A-RF sample profiles. It should be noted that cutoff dates to enter school during the time the data was collected was later than more contemporary students today. Participants completed Form TX, which contained 704 items including the 550 original MMPI items and 154 new items. Participants also completed the Biographical Information (age, religion, ethnic origin, grade level, average school grades, school activities, school problems, future plans, father's and mother's educational levels, father's and mother's occupations, number of siblings, current level arrangements) and Life Events forms (list of positive, negative, or neutral life events within the past six months) (Archer, Handel, Ben-Porath, & Tellegen, 2016).

The MMPI-A-RF normative sample is composed of 1,610 students (805 boys, 805 girls. The exclusion criteria (Cannot Say > 29; T-scores of VRIN, TRIN, L, or K > 79; or F > 89) were used to remove invalid protocols and only participants who completed all three measures (Form

TX, Biographical Information form, Life Events form) were included in the sample. The sample ranges in age from 14-18 years old (*M* = 15.54, *SD* = 1.17 for boys and *M* = 15.59, *SD* = 1.18 for girls). In terms of ethnicity, the sample is composed of 1,229 Whites (76.3%), 199 Blacks (12.4%), 46 Asians (2.9%), 46 Native Americans (2.9%), 33 Hispanics (2.0%), 41 Other (2.5%), and 16 who did not report ethnicity (1.0%; Archer et al., 2016). See Table 1 for the age, grade level, mother's education, and father's education distribution of the American MMPI-A-RF adolescent normative sample (Archer et al., 2016).

Table 1. *Distribution of Age, Grade Level, Mother's Education, and Father's Education for the American and Korean MMPI-A-RF Adolescent Normative Samples*

	American	Korean
Age		
14	366 (22.7%)	240 (19.3%)
15	438 (27.2%)	243 (19.5%)
16	427 (26.5%)	265 (21.3%)
17	292 (18.1%)	297 (23.8%)
18	87 (5.4%)	201 (16.1%)
Grade Level		
7th or Freshmen in Middle School	8 (0.5%)	2 (0.2%)
8th or Sophomore in Middle School	121 (7.5%)	187 (15.0%)
9th or Senior in Middle School	415 (25.8%)	245 (19.7%)
10th or Freshmen in High School	470 (29.2%)	270 (21.7%)
11th or Sophomore in High School	385 (23.9%)	285 (22.9%)
12th or Senior in High School	210 (13.0%)	257 (20.6%)
None Reported	1 (<0.1%)	
Mother's Education		
Graduate School	159 (9.9%)	22 (1.8%)
College Graduate	503 (31.2%)	250 (20.1%)
Some College	325 (20.2%)	
High School Graduate	476 (29.6%)	750 (60.2%)
Some High School	92 (5.7%)	
Middle School Graduate		152 (12.2%)
Grade School	18 (1.1%)	61 (4.9%)
No School		6 (0.5%)
None Reported	37 (2.3%)	5 (0.4%)

Table 1 Continued

		American	Korean
Father's Education			
	Graduate School	274 (17.0%)	82 (6.6%)
	College Graduate	531 (33.0%)	363 (29.1%)
	Some College	222 (13.8%)	
	High School Graduate	362 (22.5%)	609 (48.9%)
	Some High School	145 (9.0%)	
	Middle School Graduate		116 (9.3%)
	Grade School	29 (1.8%)	68 (5.5%)
	No School		3 (0.2%)
	None Reported	47 (2.9%)	5 (0.4%)

Korean adolescent normative sample. The Korean adolescent normative sample for the Korean MMPI-A-RF is a subset of the Korean MMPI-A normative sample. It was collected from 2003 – 2004 and gathered from middle and high school students throughout several cities in Korea: Seoul, Incheon, Kyungki, Kangwon, Daejeon, Chungnam, Chungbuk, Kwangju, Jeonnam, Jeonbuk, Jeju, Busan, Daegu, Kyungnam, and Kyungbuk. These cities were specifically chosen based on the 2000 census to ensure the heterogeneity of the sample according to geographic location, rural-urban residence, sex, and grade.

The initial sample included 1,686 Korean adolescents. However, ninety-five participants were excluded from the sample based on the exclusion criteria: Cannot Say > 0, Raw scores of VRIN > 15, TRIN < 4, TRIN > 15, or F > 34. It should be noted that the exclusion criteria for the Korean adolescent normative sample is different than the American adolescent normative sample. Studies have shown that using the American-derived validity criteria for Korean samples may exclude an excessive number of valid cases (Cheung, Song, & Zhang, 1996; Ketterer, Han, Hur, & Moon, 2010). Therefore, specific Korean exclusion criteria were created to account for these differences. After removing invalid cases, the sample consisted of 1,518 students (759

boys, 759 girls). The sample ranged in age from 12-18 years old (M = 15.41, SD = 1.73 for boys and M = 15.43, SD = 1.73 for girls). In order to be consistent with the age range of the American adolescent normative sample, participants who are under the age of 14 were excluded from this study. A total of 275 12 and 13-year-old participants were removed. The Biographical Information form and the Life Events form were also collected with this sample. The final Korean adolescent normative sample for this study is 1,246 (624 boys and 622 girls). See Table 1 for the age, grade level, mother's education, and father's education distribution of the Korean MMPI-A-RF adolescent normative sample (Han, Lim, Moon, Yook, & Kim., in press). When calculating for differences in age between Korean and American samples, Americans adolescents had a significantly higher means in ages 14 [$t(2855) = 5.14$, $p < .001$], 15 [$t(2855) = 7.55$, $p < .001$], and 16 [$t(2855) = 6.20$, $p < .001$] while Korean adolescents had a significantly higher mean for 18 year olds [$t(2855) = -6.77$, $p < .001$].

Instruments

MMPI-A-RF. Because the MMPI-A-RF was thoroughly discussed above, this section will provide a brief overview and present the psychometric properties of the assessment. The MMPI-A-RF used the structure of the MMPI-2-RF as the basis of its development. By incorporating the measure of demoralization, intercorrelations between the clinical scales decreased and allowed for each construct to be better assessed. The MMPI-A-RF uses uniform T scores to transform the raw-score distribution of the scale to preserve the positive skewness of the MMPI and to minimize the discrepancies from the original linear T-score distributions of the Clinical scales (Tellegen & Ben-Porath, 1992). The MMPI-A-RF uses the T score of 60 as a cut-off to identity clinically significant elevations. It can be hand scored, computer scored using a software program, or mailed to Pearson for scoring (Archer et al., 2016).

Psychometric properties of the MMPI-A-RF. The psychometric properties of the MMPI-A-RF were based on the analyses of the MMPI-A data sets and calculated in a variety of settings including the normative sample, and the development samples from inpatient, outpatient, correctional, and school settings. The MMPI-A-RF shows adequate reliability. Cronbach's alpha for the Validity Scales (F-r, L-r, K-r) ranged from .42 to .77 across samples, .53 to .74 in the normative sample, .47 to .76 in the outpatient sample, .43 to .77 in the inpatient sample, .49 to .76 in the correctional sample, and .42 to .75 in the school sample. Test-retest reliabilities ranged from .51 to .74 in all samples. For the VRIN-r, TRIN-r, and CRIN validity scales, alpha coefficients ranged from .24 to .60 in the normative sample, .10 to .42 in the outpatient sample, .09 to .43 in the inpatient sample, .08 to .46 in the correctional sample, and .04 to .39 in the school sample. These alpha levels were expectedly the lowest in the measure since extreme content nonresponsiveness does not occur in valid protocols (Archer et al., 2016).

For the Externalizing scales, the Cronbach's alpha coefficients and test-retest correlations were examined in the normative, outpatient, inpatient, correctional, and school samples. In the correctional sample, five of the six scales have alpha coefficients greater than .60 and ranged from .48 to .77. It is important to note that alpha coefficients are a measure of internal consistency reliability in a specific sample. Therefore, alpha coefficients are not a fixed property of a scale. For example, if a measure of psychosis were administered to college students, the measure would be expected to have low alpha coefficients because of the lack of score variability of psychosis in the college sample. In this sense, alpha coefficients for the Externalizing scales would not be expected to be high in normative or non-clinical samples but adequately reliable in appropriate settings (i.e., correctional sample). However, even with the expectedly lower alpha coefficients in the normative samples, examining the measurement

invariance in the normative sample is still warranted since the normative samples are used to generate the T scores for the tests. The Cronbach's alphas ranged from .29 to .62 in the normative sample, .41 to .74 in the outpatient sample, .42 to .78 in the inpatient sample, and .41 to .74 in the school sample. Test-retest coefficients ranged from .46 to .71 in all samples. While Cronbach's alpha coefficients under .70 do not always reflect inadequate reliability (Schmitt, 1996), among the six externalizing scales, there are four scales (NSA, CNP, SUB, AGG) whose coefficients exceed .70 (Archer et al., 2016).

Since the MMPI-A-RF was developed using a wide sample range (normative, outpatient, inpatient, etc.), the reliability and generalizability of its construct validity may decrease given the variability in samples. Standard of error of measurements are less dependent on the variability in the sample of test takers (American Educational Research Association, 2014) and are deemed adequate for the Externalizing Scales. The standard error of measurements for the Externalizing Scales are as follows: NSA (7 = normative sample; 7 = normative test-retest sample), ASA (7 = normative sample; 6 = normative test-retest sample), CNP (6-7 = normative sample; 5 = normative test-retest sample), SUB (7 = normative sample; 7 = normative test-retest sample), NPI (10 = normative sample; 6 = normative test-retest sample), and AGG (6-7 = normative sample; 6 = normative test-retest sample). It is also important to note that the standard error of measurements are relative to the full range of possible T scores on each scale (20-100) such that a standard error of measurement score of 7 is not significantly large. Extensive correlations were computed as part of the development of the measure and are present in tabular form in the manual (Archer et al., 2016).

The MMPI-A-RF scales were correlated with the MMPI-A scales to assess for construct, convergent, and divergent validity. The correlations were adequate and in the expected

directions. All scales were tested with an acute inpatient psychiatric setting (Veltri et al., 2009), residential treatment facility, (Forbey & Ben-Porath, 2003), predispositional forensic data (Handel et al., 2011), and an archival forensic sample (Veltri et al., 2009). These analyses also include extra-test variables used in the original published articles.

Korean MMPI-A-RF. The Korean MMPI-A-RF is currently being published by Kyunghee Han, Ph.D. and the research staff at Maumsarang Inc., who hold the license for the Korean versions of the MMPI measures. All of the items of the Korean MMPI-A-RF are a subset of the Korean MMPI-A. It mimics the American MMPI-A-RF in the content and number of scales. The American MMPI-A has been translated and validated into the Korean language by two individuals, fluent in Korean and English, independently translating each item (Lim & Han, 1999). Any discrepancies between these two translations were then re-evaluated and retranslated by these two translators through mutual consensus. Items were then back translated into English by another individual fluent in both Korean and English. The back translated items and the original English MMPI-A items were then examined for discrepancies by Dr. James Butcher, who is an expert in MMPI cross-cultural research and who was a member of the MMPI Restandardization Committee. Lastly, based on his review, several items were retranslated for final publication. It is projected that the Korean MMPI-A-RF will be published by the fall of 2018. Although not officially published, the psychometric properties of the Korean MMPI-A-RF have been computed and are presented below.

The psychometric properties of the Korean MMPI-A-RF were based on the analyses of the Korean MMPI-A data sets and calculated in the Korean adolescent normative and clinical sample. The Korean clinical sample totaled 237 adolescents. The Korean MMPI-A-RF Externalizing Scales showed adequate reliability. In the Korean normative sample, Cronbach's

alpha levels for males ranged from .40 to .55 and ranged from .39 to .61 for females. The standard error of measurements for the Externalizing Scales in the Korean normative sample were as follows: NSA (males = 7; females = 7), ASA (males = 7; females = 7), CNP (males = 7; females = 7), SUB (males = 7; females = 7), NPI (males = 7; females = 7), and AGG (males = 7; females = 7). In the Korean clinical sample, Cronbach's alpha levels ranged from .58 to .72. The standard error of measurements for the Externalizing Scales in the Korean clinical sample were as follows: NSA (7), ASA (7), CNP (8), SUB (5), NPI (6), and AGG (6).

Statistical Analyses

Data preparation. To prepare for the analyses, the normative data of the Korean and American MMPI-A-RF was assessed to meet statistical assumptions to perform the following analyses. In addition, both data sets were examined according to validity criteria and invalid cases were removed. Cases with any missing responses on items were also removed. Data was recoded in that a keyed response was coded as one while an unkeyed response was coded as zero. Mplus input files were created for each Externalizing Scale by creating text (.txt) files from SPSS files.

Model Specification and analysis. As noted above, the sequence of tests for multiple-group CFA follow a slightly different format when using categorical variables which is outlined in this section. A CFA model was tested separately in each group to establish that a one-factor model for each externalizing scale was acceptable in both groups. If both groups show good model fit, the first step testing for configural invariance is mostly the same for continuous and categorical variables. Both groups were analyzed simultaneously and model fit is calculated to establish configural invariance. While the thresholds and factor loadings were freely estimated in

both groups, the residual variances were fixed to 1.00 and the factor means were fixed to zero for identification purposes (Ketterer, 2010).

The second step testing for factorial invariance combines both step two and three in the sequence above where thresholds were examined instead of intercepts. The thresholds and factor loadings were constrained to equality, while the residual variances were fixed to 1.00 in one group and freely estimated in the second group, and the factor mean was fixed to zero in one group and freely estimated in the second. Constraining the thresholds and factor loadings to equality allowed for the examination of change in model fit between the factorial invariance model with the configural invariance model. In addition, fixing the residual variances and factor means to 1.00 and zero, respectively, in one group and freely estimating them in another also allowed examination of change in model fit. If there is no reduction in model fit, or if there is a nonsignificant model fit difference, factorial invariance is established and any difference between the estimated latent means can be interpreted.

If full factorial invariance is not established, partial factorial invariance will be calculated to identify noninvariant items of each scale. This is achieved by successively constraining the factor loadings and thresholds one item at a time while keeping free the remaining items. Residual variances were fixed to 1.00 in both groups and factor means were fixed to zero in both groups. Comparing these partial factorial invariance models with the configural invariance model, a noninvariant item would result in a significant model fit difference. Following Bryne, Shavelson, and Muthén's (1989) suggestion, latent means were calculated for scales that reach partial factorial invariance by simultaneously freeing the equality constraints on the factor loadings thresholds of noninvariant items while constraining the remaining items on the scale. Residual variances were fixed to 1.00 in both groups and factor means were fixed to zero in one

group and freed in the other. Due to their partial invariance, latent means were not interpreted but calculated for reference purposes.

The third and fourth steps regarding residual variance and latent mean equality are only appropriate if full factorial invariance is established in step two. Step three involves placing an additional fixed constraint on the error variances to 1.00. Residual invariance of the y* variables is established if there is no reduction in model fit when compared to the model in step two. Lastly, by placing an additional fixed constraint on the latent means to zero, latent mean invariance is established if there is no reduction in model fit when compared to the model in step two.

Goodness-of-fit indices. In order to evaluate the model fit for each scale in the Korean and American normative samples, RMSEA, CFI, and TLI were computed. These indices evaluated how well the CFA model fits or reproduces the input data. As previously discussed, good model fit is obtained when RMSEA values are close to .06 or below, and CFI and TLI values are close to .95 or above (Hu & Bentler, 1999). Guidelines suggested by Browne and Cudeck (1993) and Bentler (1990) were also considered as follows: RMSEA values less than .08 indicated adequate model fit, RMSEA values below .05 showed good model fit, and RMSEA values equal to or greater than 0.1 were rejected. In addition, CFI and TLI values lower than .90 were rejected, while values ranging from .90 to .95 may be acceptable. When comparing two models with continuous indicators, a chi-squared difference test, $\Delta\chi^2$ is used. It is calculated as the difference of χ^2 values between two models, because the difference in χ^2 value remains in the χ^2 distribution. However, for categorical indicators, the DIFFTEST in Mplus is calculated instead, since the difference in χ^2 value is not in the χ^2distribution (Brown, 2006). Due to the large sample size of each normative sample and the literature surrounding measurement

invariance testing, a *p* value of <.001 was chosen as the significance level all calculations (Wang, 2004; Wang, 2017).

CHAPTER IV

RESULTS

Descriptive Statistics

Table 2 presents the endorsement percentages in the keyed direction of the Externalizing Scales items across Korean and American groups. Higher values indicated more reported psychopathology. Higher average endorsement percentages were not calculated for significant differences, although these were not tested for statistical significance. Across cultures, both Korean and American groups reported higher average endorsement percentages in three scales each, descriptively. The Korean group indicated higher average endorsement percentages in the Negative School Attitudes (NSA), Antisocial Attitudes (ASA), and Aggression (AGG) scales than the American group while the American group showed higher endorsement percentages in the Conduct Problems (CNP), Substance Use (SUB), and Negative Peer Influence (NPI) scales than the Korean group.

Table 2. *Percentage Responding for Externalizing Scale Items in Keyed Direction, and Internal Consistency Coefficients*

Externalizing Scales	Items	American	Korean
NSA			
	29	17	**22**
	75	31	**51**
	104	**24**	18
	136 (r)	30	**32**
	195	53	**73**
	241	40	**66**
	M	32.50	**43.67**
	SD	12.63	23.14
	α	.54	.52
ASA			
	35	**64**	57
	80	38	**42**

Table 2 Continued

Externalizing Scales	Items	American	Korean
ASA			
	99	**47**	45
	171	**66**	**66**
	193	57	**72**
	219	51	**62**
	M	53.83	**57.33**
	SD	10.65	11.83
	α	.50	.41
CNP			
	14	**48**	37
	33 (r)	**29**	27
	88	**15**	3
	110	**19**	11
	127	**24**	22
	148	**12**	7
	238 (r)	**26**	14
	M	**24.71**	17.29
	SD	11.91	12.01
	α	.60	.44
SUB			
	43	14	**22**
	72	**17**	8
	166	**22**	11
	235	**23**	4
	M	**19.00**	11.25
	SD	4.24	7.72
	α	.55	.49
NPI			
	19	**19**	11
	64	25	**27**
	111	25	**43**
	146	**30**	24
	160	**30**	23
	M	**25.80**	25.60
	SD	4.55	11.48
	α	.57	.59
AGG			
	16	59	**75**
	36	15	**20**
	41	**69**	62
	130	43	**68**
	149	**28**	26
	186	21	**26**

Table 2 Continued

Externalizing Scales	Items	American	Korean
AGG			
	233	43	**57**
	240 (r)	**27**	22
	M	38.13	**44.50**
	SD	18.86	23.10
	α	.57	.55

Note. Higher percentages are bolded. NSA = Negative School Attitudes. ASA = Antisocial Attitudes. CNP = Conduct Problems. SUB = Substance Abuse. NPI = Negative Peer Influence. AGG = Aggression.

Confirmatory Factor Analysis

Table 3 presents the CFA model fit of a one-factor model for each Externalizing Scale across Korean and American groups. Based on the goodness-of-fit indices of CFI, TLI, and RMSEA, most of the scales indicated good model fit across groups except for the NSA (*TLI* = .861), ASA (*CFI* = .827; *TLI* = .712), CNP (*TLI* = .883), and NPI (*CFI* = .848; *TLI* = .695; *RMSEA* = .151) scales for the Korean group and the NPI (*CFI* = .898; *TLI* = .796; *RMSEA* = .105) scale for the American group.

Table 3. *Confirmatory Factor Analysis of Externalizing Scales Across Cultures*

	Model	χ^2	*df*	*p*	CFI	TLI	RMSEA
NSA							
	American normative	27.380	9	.001	.976	.960	.036
	Korean normative	57.290	9	<.001	.917	.861	.066
ASA							
	American Normative	23.160	9	.006	.973	.956	.031
	Korean normative	52.096	9	<.001	.827	.712	.062
CNP							
	American normative	65.855	14	<.001	.963	.945	.048
	Korean normative	43.664	14	<.001	.922	.883	.041
SUB							
	American normative	18.917	2	.001	.970	.911	.072
	Korean normative	1.705	2	.426	1.000	1.002	.000

Table 3 Continued

	Model	χ^2	*df*	*p*	CFI	TLI	RMSEA
NPI							
	American normative	94.270	5	<.001	.898	.796	.105
	Korean normative	146.907	5	<.001	.848	.695	.151
AGG							
	American normative	58.610	20	<.001	.964	.950	.035
	Korean normative	41.286	20	.004	.974	.963	.029

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. NSA = Negative School Attitudes. ASA = Antisocial Attitudes. CNP = Conduct Problems. SUB = Substance Abuse. NPI = Negative Peer Influence. AGG = Aggression.

Modification indices were examined to determine whether freeing parameters would decrease the χ^2 difference to improve the model fit. Items 19 and 146 of the NPI scale were the only items that showed a justifiable model fit improvement in both Korean ($\Delta\chi^2 = 126$) and American ($\Delta\chi^2 = 52$) groups after calculating its modification indices. After further review of items 19 and 146 of the NPI scale, it was justified that these items were highly similar in content and expressed comparable meaning. The errors of items 19 and 146 were correlated, which improved the model fit for both Korean (*CFI* = 1.000; *TLI* = 1.007; *RMSEA* = .000) and American (*CFI* = 1.000; *TLI* = 1.007; *RMSEA* = .000) groups to be acceptable for further analysis.

While Browne and Cudeck (1993) and Bentler (1990) recommend using cutoffs for goodness-of-fit indices CFI and TLI of lower than .90, researchers also report controversy in setting strict reliance on cutoff guidelines (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). For the NSA and CNP scales in the Korean group, the TLI fit indices were lower than the suggested .90 cutoff. Kenny (2015) reported that because the CFI and TLI are highly correlated, only one of the two fit indices should be reported. After further examination, the NSA (*CFI* = .917; *RMSEA* = .066) and CNP (*CFI* = .922; *RMSEA* = .041) scale of the

Korean group were deemed acceptable for continued analysis as evidenced by the good model fit indicated by both the CFI and RMSEA fit indices.

Table 4 presents the factor loadings of each scale. The American group had higher average factor loadings on nearly all of the scales: NSA (American λ_{mean} = .54, Korean λ_{mean} = .53), ASA (American λ_{mean} = .49, Korean λ_{mean} = .42), CNP (American λ_{mean} = .58, Korean λ_{mean} = .50), SUB (American λ_{mean} = .68, Korean (λ_{mean} = .63), AGG (American λ_{mean} = .49, Korean λ_{mean} = .48). For the NPI scale, the Korean group (λ_{mean} = .67) had higher average factor loadings than the American group (λ_{mean} = .61).

Table 4. *Standardized Factor Loadings and Thresholds for Externalizing Scales by Culture*

		American		Korean	
Externalizing Scales	Items	FL	TH	FL	TH
NSA					
	29	.66	.96	.74	.76
	75	.72	.49	.42	-.02
	104	.62	.72	.51	.92
	136 (r)	.46	.53	.61	.47
	195	.38	-.08	.57	-.61
	241	.37	.24	.35	-.41
	M	**.54**	.48	.53	.19
	SD	.15	.36	.14	.63
ASA					
	35	.58	-.36	.50	-.19
	80	.50	.32	.41	.22
	99	.37	.07	.36	.13
	171	.46	-.41	.35	-.42
	193	.49	-.17	.58	-.58
	219	.52	-.02	.31	-.30
	M	**.49**	-.10	.42	-.19
	SD	.07	.28	.10	.31
CNP					
	14	.45	.06	.43	.35
	33 (r)	.61	.55	.24	.61
	88	.79	1.06	.68	1.90
	110	.52	.89	.60	1.25
	127	.83	.70	.80	.77

Table 4 Continued

		American		Korean	
Externalizing Scales	Items	FL	TH	FL	TH
CNP					
	148	.47	1.20	.25	1.45
	238 (r)	.37	.65	.49	1.07
	M	**.58**	.73	.50	1.06
	SD	.18	.38	.21	.53
SUB					
	43	.70	1.10	.56	.78
	72	.70	.97	.79	1.38
	166	.56	.76	.96	1.21
	235	.67	.74	.21	1.76
	M	**.68**	.89	.63	1.28
	SD	.07	.17	.32	.41
NPI					
	19	.70	.89	.82	1.25
	64	.60	.69	.66	.62
	111	.61	.69	.54	.17
	146	.66	.52	.67	.70
	160	.49	.53	.67	.74
	M	.61	.66	**.67**	.70
	SD	.08	.15	.10	.38
AGG					
	16	.39	-.22	.47	-.69
	36	.41	1.03	.36	.84
	41	.41	-.51	.70	-.31
	130	.71	.18	.67	-.46
	149	.69	.58	.53	.66
	186	.39	.79	.58	.64
	233	.56	.17	.40	-.17
	240 (r)	.38	.61	.12	.78
	M	**.49**	.33	.48	.16
	SD	.14	.52	.19	.63

Note. Higher percentages are bolded. NSA = Negative School Attitudes. ASA = Antisocial Attitudes. CNP = Conduct Problems. SUB = Substance Abuse. NPI = Negative Peer Influence. AGG = Aggression.

Measurement Invariance Tests

Measurement invariance tests were conducted for each scale across cultures, except for the ASA scale since the Korean group did not indicate good model fit and using modification

indices to improve model fit was not justified. The configural model fit was acceptable for all tested scales. However, when testing for factorial invariance, the model fit difference between the factorial model and the configural model was significant for all tested scales, indicating that full factorial invariance was not reached for all tested scales. Partial factorial invariance was calculated and met for all tested scales except for the SUB scale, which only met configural invariance and failed to converge for partial factorial invariance (see Table 7). For scales that reached partial factorial invariance, latent means were calculated and compared between groups. All of the latent means calculated did not differ significantly between Korean and American groups. The NSA (Z = -.002, p = .980), CNP (Z = -.051, p = .589), and AGG (Z = -.073, p = .167) latent means were nonsignificantly lower in Korean adolescents compared to American adolescents while the NPI (Z = .136, p = .407) latent mean was nonsignificantly higher in Korean adolescents compared to American adolescents. Noninvariant items were identified for each scale across cultures.

For the NSA scale, three noninvariant items (75, 195, 241) were identified across Korean and American groups (see Table 5). For the CNP scale, five noninvariant items (14, 33, 88, 110, 238) were found (see Table 6). For the NPI scale, there were three noninvariant items (19, 111, 160) (see Table 8). For the AGG scale, four noninvariant items (16, 130, 233, 240) were identified (see Table 9).

Table 5. *Fit Indices for Invariance Models across Normative Samples for Negative School Attitudes*

	χ^2	*df*	RMSEA	CFI	TLI	Model Comparison	$\Delta\chi^2$ (df)	*p*
Model 1: Configural Invariance	85.028	18	.051	.950	.917			
Model 2a: Factorial Invariance	176.701	23	.068	.885	.850	2a vs. 1	81.357 (5)	<.001
Model 2b: Partial Factorial Invariance								
29	87.176	20	.048	.950	.925	29 vs. 1	3.047 (2)	.218
75	257.561	20	.091	.823	.734	75 vs. 1	140.414 (2)	<.001
104	99.480	20	.053	.941	.911	104 vs. 1	13.659 (2)	.001
136	83.461	20	.047	.953	.929	136 vs. 1	.638 (2)	.727
195	239.379	20	.088	.836	.755	195 vs. 1	124.145 (2)	<.001
241	306.341	20	.100	.786	.680	241 vs. 1	170.079 (2)	<.001

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. Significant *p* value = <.001. Noninvariant items are bolded.

Table 6. *Fit Indices for Invariance Models across Normative Samples for Conduct Problems*

	χ^2	*df*	RMSEA	CFI	TLI	Model Comparison	$\Delta\chi^2$ (df)	*p*
Model 1: Configural Invariance	108.705	28	.045	.955	.933			
Model 2a: Factorial Invariance	139.260	34	.047	.941	.928	2a vs. 1	30.815 (6)	<.001
Model 2b: Partial Factorial Invariance								
14	159.542	30	.055	.928	.899	14 vs. 1	37.524 (2)	<.001
33	127.808	30	.048	.946	.924	33 vs. 1	16.328 (2)	<.001
88	141.497	30	.051	.938	.913	88 vs. 1	26.159 (2)	<.001
110	146.003	30	.052	.935	.910	110 vs. 1	29.336 (2)	<.001
127	106.665	30	.042	.957	.940	127 vs. 1	0.266 (2)	.875
148	120.230	30	.046	.950	.930	148 vs. 1	11.111 (2)	.004
238	178.759	30	.059	.917	.884	238 vs. 1	50.031 (2)	<.001

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. Significant *p* value = <.001. Noninvariant items are bolded.

Table 7. *Fit Indices for Invariance Models across Normative Samples for Substance Abuse*

	χ^2	*df*	RMSEA	CFI	TLI	Model Comparison	$\Delta\chi^2$ (df)	*p*
Model 1: Configural Invariance	17.970	4	.049	.987	.961			
Model 2: Factorial Invariance	89.763	7	.091	.924	.869	2a vs. 1	68.191 (3)	<.001

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. Significant *p* value = <.001. Noninvariant items are bolded.

Table 8. *Fit Indices for Invariance Models across Normative Samples for Negative Peer Influences*

	χ^2	*df*	RMSEA	CFI	TLI	Model Comparison	$\Delta\chi^2$ (df)	*p*
Model 1: Configural Invariance	2.473	6	.000	1.00	1.01			
Model 2a: Factorial Invariance	129.393	10	.091	.934	.868	2a vs. 1	100.082 (4)	<.001
Model 2b: Partial Factorial Invariance								
19	37.426	8	.051	.984	.959	19 vs. 1	23.835 (2)	<.001
64	3.795	8	.000	1.00	1.01	64 vs. 1	1.003 (2)	.606
111	92.197	8	.086	.953	.883	111 vs. 1	60.663 (2)	<.001
146	16.133	8	.027	.995	.989	146 vs. 1	9.770 (2)	.008
160	37.391	8	.051	.984	.959	160 vs. 1	23.818 (2)	<.001

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. Significant *p* value = <.001. Noninvariant items are bolded.

Table 9. *Fit Indices for Invariance Models across Normative Samples for Aggression*

	χ^2	*df*	RMSEA	CFI	TLI	Model Comparison	$\Delta\chi^2$ (df)	*p*
Model 1: Configural Invariance	99.973	40	.032	.968	.955			
Model 2a: Factorial Invariance	408.513	47	.073	.806	.769	2a vs. 1	226.389 (7)	<.001
Model 2b: Partial Factorial Invariance								
16	205.498	42	.052	.912	.883	16 vs. 1a	68.618 (2)	<.001
36	111.529	42	.034	.963	.950	36 vs. 1a	9.560 (2)	.008
41	113.680	42	.035	.961	.949	41 vs. 1a	11.164 (2)	.004
130	311.497	42	.067	855	.807	130 vs. 1a	160.967 (2)	<.001
149	112.115	42	.034	.962	.950	149 vs. 1a	10.303 (2)	.006
186	102.674	42	.032	.967	957	186 vs. 1a	3.882 (2)	.144
233	177.806	42	.048	.927	.903	233 vs. 1a	55.473 (2)	<.001
240	139.006	42	.040	.948	.931	240 vs. 1a	25.931 (2)	<.001

Notes. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. Significant *p* value = <.001. Noninvariant items are bolded.

CHAPTER V

DISCUSSION

The present study examined the measurement invariance of the MMPI-A-RF Externalizing Scales across Korean and American adolescent normative samples. The hypothesis of this study was partially supported: All Externalizing Scales met partial factorial invariance, indicating a one-factor model, except for the Antisocial Attitudes (ASA) scale, which did not indicate a one-factor model in the Korean sample, and the Substance Abuse (SUB) scale, which only reached configural invariance. None of the scales had reached full factorial invariance. For scales that reached configural invariance, these results suggest that the same items are associated with the same latent factor in each group. In addition, the invariant items of each scale that reached partial factorial invariance have the same meaning and bear the same weight across cultures. In other words, for those invariant items, for any given level of the latent factor, the expected score for an item was equivalent across Korean and American samples. However, the results also showed that there were noninvariant items within each scale. Noninvariant items did not indicate the same meaning or bear the same weight across culture and for any given level of the latent factor, the expected score for an item was not equivalent across Korean and American samples. Noninvariant items are further discussed by examining the cultural differences between Korean and American adolescents.

While the Korean and American normative adult samples for the MMPI-2 did not indicate a one-factor model in the RC Scales that encompass the Externalizing Scales (RC4, RC9), most of the Externalizing Scales in the present study indicated a one-factor model (Ketterer, 2010). In addition, while Wang (2014) found partial scalar invariance of the MMPI-2 RC4 scale four-factor model (School Problems, Substance Abuse, Family Problems, Violation of

Social Norms) between Korean and American clinical adult samples, all but two MMPI-A-RF Externalizing Scales reached partial factorial invariance in this study. Further, although Wang (2017) found partial scalar invariance in the MMPI-2-RF Externalizing Specific Problems Scales between Korean and American normative and clinical adult samples except for the JCP scale which received full scalar invariance; most scales in this study reached partial factorial invariance but none had reached full factorial invariance.

Although the ASA scale indicated a one-factor model in the American sample, it did not show a one-factor model in the Korean sample. As the ASA scale did not fit a one-factor model in Korean adolescents, no further testing could be calculated. For both Korean and American adolescents, the literature indicates similar risk factors for juvenile delinquency including negative family dynamics, deviant peers, and comorbid mental disorders such as depression and substance use (Blum & Libbey, 2004; Kim & Kim, 2002; Lee & Jun, 2009; Lee, Onifade, Teasley, & Noel, 2012). The multidimensionality of the ASA scale in Korean adolescents shows that while influences for antisocial behaviors and attitudes may develop from cross-cultural characteristics and circumstance, the construct of delinquency in Korean adolescents may be conceptualized from a different cultural lens when compared to American adolescents. These results appear to be consistent with Ketterer (2010) and the objective of Wang (2014) in their efforts to investigate a one-factor model in Korean samples. Noninvariant items of each scale including the ASA scale may be further explained through these varying perspectives.

The SUB scale is the only scale that reached configural invariance but did not obtain partial factorial invariance. These results may be best captured by the culture related to drinking and substance use in Korea. Drinking alcohol in the Korean community is heavily tolerated and seen as a social and bonding experience in interpersonal relationships. Men are encouraged to

drink with co-workers and friends and sometimes their drinking capacities indicate social competence (Kwon-Ahn, 2001). While the prevalence rates of alcohol consumption in Korean high school students (44.5%) were comparable to American high school students (48.6%), marijuana use for Korean and American students was not (Johnston, O'Malley, & Bachman, 2003; Kim & Kim, 2002). There are significantly higher rates of marijuana use in American high school students (37%) compared to Korean students (0.22%) (U.S. Department of Health & Human Services, 2015; Park & Kim, 2016). While prevalence rates for alcohol consumption between Korean and American adolescents were comparable, their perspective on excessive use and what they deem as a problem appear to have varying cultural meanings. However, although the SUB scale did not meet partial factorial invariance, White (2017) found that the SUB scale had high correlations with both therapist and self-report substance abuse criteria in Korean normative and clinical samples, accurately predicting substance abuse in Korean adolescents. This shows that while the construct of substance use may vary across Korean and American cultures, the SUB scale does show strong external validity and predictive capability for substance abuse in Korean adolescents.

For Negative Attitudes School (NSA) scale, the three noninvariant items (75, 195, 241) were related to negative attitudes towards school including finding academics boring, feeling sleepy, and preferring to be with friends to avoid school activities. Although friends, teachers, and parents were all found to be associated with school satisfaction in Korean adolescents, Kim and Kim (2013) showed that the teacher-student relationship was the most important predictor for school satisfaction for high school students in Korea. Although the Korean sample had a higher average level of endorsement for these items, items based on friendships and negative school attitudes may be interpreted differently in Korean society due to the value placed on

education and the respected role of an educator. Sleepiness in school may also be conceptualized from another point of view in Korea due to certain consequences for inappropriate behaviors in school. Direct corporal punishment, which involves any type of hitting, and indirect corporal punishment, like holding stress positions, have been part of Korea's school disciplinary system for many decades (Marquez, 2015). Although all direct corporal punishment was banned by the Seoul Metropolitan Office of Education in 2010, Korean students still experience corporal punishment in school (Lee, 2016). Due to the strict environment and consequences, sleepiness in school may be interpreted differently for Korean students compared to American students. Lastly, when it comes to school activities in Korea, Kim and Kim (2013) found that there were no associations with school activities and school satisfaction.

The Conduct Problems (CNP) scale had the most noninvariant items (14, 33, 88, 110, 238) of all the scales (5 out of 7) and was the only scale where American adolescents had higher item endorsement compared to Korean adolescents for all of its items. The noninvariant items of this scale were related to negative conduct behaviors at school and with the law. One of the reasons why negative conduct behaviors at school may be interpreted differently in Korean adolescents is due to the extreme pressures of excelling in school and passing the entrance examinations (Korea University Scholastic Ability Test) for admission to a high ranked university, which may influence the need to act and present appropriately in school (Kim & Kim, 2008). Chung, Kim, Lee, Kwon, and Lee (1993) reported that Korean twelfth graders spent as much as 14 to 18 hours per day studying and preparing for this exam. Won (1989) also reported that Korean high school students were involved with school work during 47% of the random time samples, compared to 25 to 29% for American adolescents. Korean parents also experience these educational pressures. Many parents feel obligated to earn additional income in order for

their children to receive extracurricular education, which may add to the conflict in parent-child relationships and psychological stress. However, in addition, these stressors may be a significant contributing factor to the rise in school violence and bullying in Korean children and adolescents (Kim & Kim, 2008). Further, for conduct problems with the law, while physical assault in surveyed Korean students decreased from 19% in 1999 to 12% in 2001, the rate of cyber crime is significantly increasing (Kim & Kim, 2002). In 2002, the most frequent types of aggressive and violent behavior in Korean adolescents were cyber terror (26.9%), extortion of money or articles (26.4%), threatening behavior or intimidation (22.9%), physical assault (21.4%) and annoyance (17.6%) (Kim & Kim, 2008). These trends indicate that negative conduct problems may be shifting from violent behavior in the real world to deviant behaviors in cyber space, which may influence their perception of conduct problems.

For the Negative Peer Influence (NPI) scale, noninvariant items (19, 111, 160) were related to parental approval of peers, peer deviance, and the influence of negative peers. In both Korean and American adolescents, negative peer influence was one of the strongest predictors of delinquent behavior (Henry et al., 2001; Lee et al., 2012; Quinsey, Skilling, Lalumiere, & Craig, 2004). In addition, much of the literature pointed to parenting styles as an important factor in the development of delinquent peers, which in turn contribute to the formation of externalizing behaviors. Behavioral control is a dimension of parenting that refers to the level of monitoring and limit setting for children and adolescents. Effective behavioral control has been associated with lower levels of externalizing behaviors as well as a decrease in the number of delinquent peers (Gray & Steinberg, 1999; Dishion, Bullock, and Granic, 2002). Psychological control is another dimension that can be characterized as the manipulation and guilt induction used by parents (Barber, Olsen, & Shagle, 1994). Stone, Buehler, and Barber (2002) report that

psychological control in parenting can increase the risk of externalizing and internalizing behaviors in adolescents. Due to the influences of Confucianism within their society, Korean parents may show higher levels of control, lower levels of parental warmth, and stronger disapproval of delinquency than American parents (Lee, Bell, & Watson, 2007). However, lower levels of parental warmth are associated with less parental knowledge for both Korean and American parents, such as an adolescent being less likely to share accurate information about his or her whereabouts, peers, and activities (Fletcher, Steinberg, & Williams-Wheeler, 2004; Son & Choi, 2013). Diana Baumrind (1967) identified three parenting styles: permissive, authoritative, and authoritarian. Permissive parenting can be defined as non-controlling, non-demanding, and warm. Authoritative parents are controlling and demanding but also warm and receptive. Authoritarian parenting is more detached, controlling, and less warm compared to permissive and authoritative parenting. Korean parents tend to be stricter and less expressive in showing affection, however, their parenting styles are not consistently related to negative youth outcomes when compared to Western authoritarian parenting (Choi, Kim, Kim, & Park, 2013). Noninvariant items of the NPI scale may have been influenced by these differences in parenting styles for Korean and American culture groups.

The noninvariant items (16, 130, 233, 240) of the Aggression (AGG) scale are related to the endorsement of acts of violence, vengeance, and aggression towards others. Difference in meaning between Korean and American adolescents for these items may be associated with the shift towards cyber related crimes of aggression in Korean adolescents, previously discussed (Kim & Kim, 2008). Many of the noninvariant items of this scale are connected to physical assaults of aggression which has been reported to be decreasing in Korean adolescents (Kim & Kim, 2002). Depression and substance use are additional factors associated with aggression

during adolescence (Kim & Lee, 2008). The Korean Ministry of Health and Welfare (2012) reported higher rates of depression in Korean adolescents (30.5%) than in American adolescents (18.6%; Kessler et al. 2012), which may contribute to the need to express emotions through varying outlets. Due to the culture surrounding substances in Korea, the high level of alcohol consumption may also add to the cultural influence of aggression in Korean adolescents. It appears that many of the contributing factors for noninvariant items including modernization, the education system, parenting, and comorbid symptoms in Korean adolescents can be intertwined, shaping the understanding of how Korean and American adolescents may view and interpret externalizing behaviors. Further research is warranted on these cultural differences and how they may impact cross-cultural psychological testing and interpretation.

Strengths, Limitations, and Future Directions

The present study is the first to examine the measurement invariance of the MMPI-A-RF across Korean and American adolescent normative samples. The development of measurement invariance testing has benefited cross-cultural researchers to ensure construct comparability and infer meaningful differences between groups. Using the original developmental samples from both the American and Korean MMPI-A-RF, this study adds to the literature of cross-cultural MMPI research and contributes to the needed investigations and discussions on the cultural validity and implications of the MMPI-A-RF Externalizing Scales.

Several limitations of this study should be noted. Although both samples used in this study were represented and collected to match the census of the country at the time, the Korean sample was collected between 2003 – 2004 to match the 2000 Korean Census in geographic location, rural-urban residence, sex, and grade. The American sample was collected in the late 1980s, early 1990s to match the American adolescent census at that time in ethnicity, geography,

and rural-urban residence. A limitation of this study is that the constructs that represent the Externalizing Scales for the MMPI-A-RF may have changed within each culture when taking into account the approximated 10-15 year gap between samples. However, in order to account for the change in constructs over time, the validation of the American MMPI-A-RF included calculations of the mean distribution of T-scores in contemporary MMPI-A scores. A mean T-score of 50 was found in contemporary scores, indicating that the American adolescent normative sample was still comparable to today's adolescents (Archer et al., 2016). Further, when accounting for age differences between samples, it was found that the American sample had significantly higher means for ages 14, 15, and 16, while the Korean sample had significantly higher means for 18 year olds. The differences in age may also have affected the invariance testing due to the inequality in age representation.

Second, due to the assumptions of assessment development and cross-cultural research, there are several reasons why noninvariance may arise (Chen, 2008). First, because of differences in cultural values across groups, the definitions and meanings behind certain concepts may be better suited in one culture compared to another. In this study, given the intricacies of individualistic and collectivistic cultures, constructs of externalizing behaviors may have had more individual or familial influences that impacted the meaning and interpretation of items. Second, translations of items between languages can cause noninvariance due to variations in idiomatic expressions. Although the Korean MMPI-A-RF was back translated, potential issues in translations may be present. Third, different populations may have varying response styles. For example, participants from the U.S. are more likely to use the extreme ends of a response scale, whereas participants from China tend to use the middle points (Chen, Lee, & Stevenson,

1995; Hui & Triandis, 1985). Issues regarding cross-cultural measurement invariance testing is the second limitation of this study.

The third limitation of this study also revolves around assessing noninvariance between cultures when analyzing group comparisons. Millsap and Kwok (2004) reported four different directions when it came to dealing with noninvariant scales. The first choice would be to eliminate the noninvariant items of each scale for further analysis. However, this would result in creating a new scale for analysis and the potential for skewing the intended construct of the scale. The second option is to keep all invariant and noninvariant items in each scale and continue with further analysis. This approach assumes that the noninvariant items of the scale would create minimal bias towards any group comparisons, which is an assumption that lacks confidence in the literature (Chen, 2008). The third option is to calculate partial measurement invariance and to interpret any further results based on constraining noninvariant items to be equal across groups. Further measurement invariance analyses based on this approach raise issues based on the validity of the altered construct and whether the interpretations of the partially invariant scales are meaningful. Additional concerns are raised when a high proportion of the items of the scale are noninvariant or when there are a few items in the scale to begin with. Lastly, when noninvariance is found, the fourth choice is to avoid making any group comparisons at all. While there are different methods to dealing with noninvariance, this study followed a combination of the third and fourth approaches by calculating partial measurement invariance and discussing potential meaningful cultural differences to explain the noninvariance found.

Further research is warranted in examining the measurement invariance of the MMPI-A-RF with the Korean and American adolescent population. Due to the multidimensionality of the

ASA, it is recommended to calculate an Exploratory Factor Analysis (EFA) to investigate whether the ASA would be better represented in a multiple-factor model rather than a one-factor model in Korean samples. Since this study examined only the normative samples between cultures, investigating clinical samples and analyses based on gender would provide insight on the relationship and validity of the Korean MMPI-A-RF. As this study only examined the Externalizing Scales of the MMPI-A-RF, it would be beneficial to analyze how other scales relate to the results of this study, providing further cultural implications and more information on the psychometric properties of the MMPI-A-RF. Due to the discrepancies in age distribution between Korean and American samples, it is also recommended to randomly select Korean cases to match the American age distribution in order to examine whether any discrepancies in age had altered the measurement invariance findings. In addition, gathering data sets that are more comparable in time may allow researchers to see how any changes in culture and behavior have impacted measures of personality and psychopathology. Finally, future directions of research should be placed on the varying aspects of construct formation and interpretation between cultures, including the Korean and American population. Cross-cultural measurement invariance testing and the challenges found within it will continue to lead to the further advancement of psychological assessment and the necessary steps in understanding the intricacies of cultural phenomenon.

BIBLIOGRAPHY

Alperin, J. J., Archer, R. P., & Coates, G. D. (1996). Development and effects of an MMPI-A K-correction procedure. *Journal of Personality Assessment, 67,* 155-168.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology scale (F_p). *Psychological Assessment, 7,* 424-431.

Archer, R. P. (1984). Use of the MMPI with adolescents: A review of salient issues. *Clinical Psychology Review, 41,* 241-251.

Archer, R. P. (1987). *Using the MMPI with adolescents.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Archer, R. P. (2005). MMPI-A: *Assessing adolescent psychopathology* (3rd ed.). Mahwah, NJ. Lawrence Erlbaum Associates.

Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment, 7,* 227-235.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87,* 84-94.

Archer, R. P., Handel, R. W., & Lynch, K. D. (2001). The effectiveness of the MMPI-A items in discriminating between normative and clinical samples. *Journal of Personality Assessment, 77,* 420-435.

Archer, R. P., Handel, R. W., Ben-Porath, Y. S., & Tellegen, A. (2016). Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF). Minneapolis, MN: University of Minnesota Press.

Archer, R. P., Handel, R. W., Ben-Porath, Y. S., & Tellegen, A. (2016). *MMPI-A-RF (Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form): Administration, scoring, interpretation, and technical manual.* Minneapolis, MN: University of Minnesota Press.

Archer, R. P., Stolberg, A. L., Gordon, R. A., & Goldman, W. R. (1986). Parent and child MMPI responses: Characteristics among families with adolescents in inpatient and outpatient settings. *Journal of Abnormal Child Psychology, 14,* 181-190.

Barber, B. K., Olsen, J. E., Shagle, S. C. (1994). Associations between parental psychological and behavioral control and youth internalized and externalized behaviors. *Child Development, 65,* 1120-1136.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology, 4,* 1-103.

Ben-Porath, Y. S., & Sherwood, N. E. (1993). *The MMPI-2 Content Component Scales: Development, psychometric characteristics, and clinical application*. 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.

Blum, R. W., & Libbey, H. P. (2004). School connectedness: Strengthening health and education outcomes for teenagers. *Journal of School Health, 74*, 229–299.

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

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A. & Long, J. S. (Eds.) *Testing Structural Equation Models* (pp. 136–162). Beverly Hills, CA: Sage

Bryson, J. H. (2017). Recent Economic Developments in South Korea. *Economics Groups*. Retrieved from http://www.wellsfargo.com

Butcher, J. N. (1996). *International adaptations of the MMPI–2: Research and clinical applications*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 25-50). Hillsdale, NJ: LEA.

Butcher, J. N., & Pancheri, P. (1976). *A handbook of cross-national MMPI research.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A, & Kaemmer, B. (1989).The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): *Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A): Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Bryne, 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.

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, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*, 170-175.

Cheung, B., Lee, C., & Jin, W. (1963). MMPI manual. Seoul: Korean Testing Center.

Cheung, F. M., Song, W. Z., & Zhang, J. X. (1996). The Chinese MMPI-2: Research and applications in Hong Kong and the People's Republic of China. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2: A handbook of research and applications* (pp. 137-161). Minneapolis, MN: University of Minnesota Press.

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

Choi, Y., Kim, Y. S., Kim, S. Y., & Park, I. J. K. (2013). Is Asian American parenting controlling and harsh? Empirical testing of relationships between Korean American and Western parenting measures. *Asian American Journal of Psychology, 4*, 19-29.

Chung, B., Kim, H., Lee, S., Kwon, K., & Lee, J. (1993). *Restoring Korean education from the bondage of entrance examination education.* Nanam Publication, Seoul.

Chung, J. J., Weed, N. C., & Han, K. (2006). Evaluating cross-cultural equivalence of the Korean MMPI-2 via bilingual test-retest. *International Journal of Intercultural Relations, 30*, 531-543.

Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 28*, 367–394.

Dahlstrom, W. G., & Welsh, G. S. (1960). *An MMPI handbook: A guide to use in clinical practice and research.* Minneapolis, MN: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Vol. I. Clinical Interpretation.* Minneapolis, MN: University of Minnesota 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.

Dishion, T. J., Bullock, B. M., & Granic, I. (2002). Pragmatism in modeling peer influence: Dynamics, outcomes, and change processes. *Development and Psychopathology, 14*, 969-981.

Dorius, C. J., Bahr, S. J., Hoffmann, J. P., & Harmon, E. L. (2004). Parenting practices as moderators of the relationship between peers and adolescent marijuana use. *Journal of Marriage and Family, 66*, 163–178.

Drake, L. E. (1946). A social I.E. scale for the MMPI. *Journal of Applied Psychology, 30*, 51-54.

Eichman, W. J. (1961). Replicated factors on the MMPI with female NP patients. *Journal of Consulting Psychology, 25*, 55-60.

Fletcher, A. C., Steinberg, L., & Williams-Wheeler, M. (2004). Parental influences on adolescent problem behavior: Revisiting Stattin and Kerr. *Child Development, 75*, 781-796.

Forbey, J. D., & Ben-Porath, Y. S. (2003). Incremental validity of the MMPI-A Content scales in a residential treatment facility. *Assessment, 10,* 191-202.

Furstenberg, F. F., Jr., Cook, T. D., Eccles, J., Elder, G. H., Jr., & Sameroff, A. (Eds.). (1999). *Managing to make it: Urban families and adolescent success.* Chicago: University of Chicago Press.

Gray, M. R., & Steinberg, L. (1999). Unpacking authoritative parenting: Reassessing a multidimensional construct. *Journal of Marriage and the Family, 61,* 574-587.

Gorman-Smith, D., Tolan, P. H., & Henry, D. B. (2000). A developmental– ecological model of the relation of family functioning to patterns of delinquency. *Journal of Quantitative Criminology, 16,* 169–198.

Graham, J. R. (2012). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.

Han, K. (1990). International applications of the MMPI-2: Translation issues and procedures. In: J. N. Butcher (Chair), Korean Translation of the MMPI-2. Paper presented at the 25th Annual Symposium on Recent Developments in the Use of the MMPI (MMPI-2), Minneapolis, MN.

Han, K. (1993). *The use of the MMPI-2 in Korea: Inventory adaptation, equivalence evaluation, and initial validation* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis.

Han, K., & Lim, J. (2001, March). *The use of Korean MMPI–2 with Korean psychiatric samples: Preliminary investigation.* Paper presented at the 36th annual symposium on the recent development of the MMPI– 2/MMPI–A, St. Petersburg, FL.

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

Han, K., Lim, J., Moon, K., Yook, K., & Kim, J. (in press). *Korean MMPI–A-RF user manual.* Seoul, Korea: Maumsarang Inc.

Han, K., Park, H. I., Weed, N. C., Lim, J., Johnson, A., & Joles, C. (2013). Gender differences on the MMPI across American and Korean adult and adolescent normative samples. *Journal of Personality Assessment, 95,* 197-206.

Hand, C. G., Archer, R. P., Handel, R. W., & Forbey, J. D. (2007). The classification accuracy of the Minnesota Multiphasic Personality Inventory – Adolescent: Effects of modifying the normative sample. *Assessment, 14,* 80-85.

Handel, R. W., Archer, R. P., Elkins, D. E., Mason, J. A., & Simonds-Bisbee, E. C. (2011). Psychometric properties of the Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A) Clinical, Content, and Supplementary scales in a forensic sample. *Journal of Personality Assessment, 93,* 566-581.

Harkness, A. R., McNulty, J. L., Ben-Porath, Y.S., & Graham, J. R. (2002). *MMPI-2 Personality Psychopathology Five (PSY-5) scales: Gaining an overview for case conceptualization and treatment planning.* Minneapolis, MN: University of Minnesota Press.

Harrison, P. L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment, 6,* 188-198.

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

Hawley, P. H. (2003). Prosocial and coercive configurations of resource control in early adolescence: A case for the well-adapted Machiavellian. *Merrill-Palmer Quarterly, 49,* 279–309.

Hayduk, L., Cummings, G. G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two three – Testing the theory in structural equation models! *Personality and Individual Differences, 42,* 841-50.

Henry, D. B., Tolan, P. H., & Gorman-Smith, D. (2001). Longitudinal family and peer group effects on violence and nonviolent delinquency. *Journal of Clinical Child Psychology, 30,* 172–186.

Hilts, D., & Moore, J. M., Jr. (2003). Normal range MMPI-A profiles among psychiatric inpatients. *Assessment, 10,* 266-272.

Hirschi, T. (1969). *Causes of delinquency*. Berkeley: University of California Press.

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

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology, 16,*

Janus, M. D., Tolbert, H., Calestro, K., & Toepfer, S. (1996). Clinical accuracy ratings of MMPI approaches for adolescents: Adding ten years and the MMPI-A. *Journal of Personality Assessment, 67,* 364-383.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (2003). *Monitoring the future national survey results on drug use, 1975–2002: Vol. 1. Secondary school students*. (NIH Pub. No. 03–5375). Bethesda, MD: National Institute of Drug Abuse.

Kessler, R. C., Avenevoli, K. A., McLaughlin, J., Greif Green, M. D., Lakoma, M., Petukhova,

D. S., & Merkangas, K. R. (2012). Lifetime co-morbidity of DSM-IV disorders in the US National comorbidity survey replication adolescent supplement (NCS-A). Psychological Medicine, 42, 1997–2010.

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)

Ketterer, H. L., Han, K., Hur, J., & Moon, K. (2010). Development and validation of culture-specific variable response inconsistency and true response inconsistency scales for use with the Korean MMPI-2. *Psychological Assessment, 22,* 504-519.

Kim, B. S., & Chung, E. H. (2004). The relationships of aggression, perceived social support and cognitive abilities of interpersonal problem solving of adolescents among addiction to internet chatting. *The Korea Journal of Youth Counseling, 12,* 91-99.

Kim, D.H., & Kim, J. H. (2013). Social relations and school life satisfaction in South Korea. *Social Indicators Research, 112,* 105-127.

Kim, H. S., & Kim, H. S. (2002). The influence of alcohol drinking and substance abuse on delinquent behavior among Korean adolescents. *Journal of Korean Neuropsychiatry Association, 41,* 472–485.

Kim, H. S., & Kim, H. S. (2008). *Juvenile delinquency and youth crime*. New York, NY: Nova.

Kim, H. S., & Lee, S. H. (2008). A study on the relations among anger expression mode, depression, and aggression in vocational high school students. *Journal of Korean Academy of Community Health Nursing, 19,* 270-279.

Kim, J. W., & Kim, H. J. (2007). Relationships among children's aggression, temperament, home environment, and school adjustment. *The Journal of Child Education, 16,* 85-93.

Kim, Y., Kim, J., Kim, J., Noh, M., Shin, D., Yum, T., & Oh, S. (1989). MMPI manual. Seoul: Korea Guidance.

Klinge, V., & Strauss, M. E. (1976). Effects of scoring norms on adolescent psychiatric patients' MMPI profiles. *Journal of Personality Assessment, 40,* 13-17.

Klinge, V., Lachar, D., Grisell, J., & Berman, W. (1978). Effects of scoring norms on adolescent psychiatric drug users' and nonusers' MMPI profiles. *Adolescence, 13,* 1-11.

Korea Ministry of Health and Welfare (2012). The eighth online survey statistics of adolescents' health. Retrieved from http://www.mohw.go.kr

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.

Le, T. N., & Stockdale, G. D. (2005). Individualism, collectivism, and delinquency in Asian American adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 681-691.

Lee, B. E. (1997). *A study for the health education curriculum in high school (*Unpublished master's thesis). Ewha Woman's University, Seoul.

Lee, J. M., Bell, N. J., & Watson, W. (2007). Situating alcohol use and delinquency within developmental and societal context: The case of Korean youth. *Journal of Adolescence, 30,* 835-851.

Lee. H. J. (2016). 20% of Seoul students still face corporal punishment: report. *The Korea Herald*. Retrieved from http://www.koreaherald.com

Lee, J., Onifade, E., Teasley, M., Noel, L. (2012). The effects of risk and protective factors on juvenile delinquency in South Korea, *Journal of Ethnicity in Criminal Justice, 10,* 316-329.

Lee, S. S., & Jun, S. H. (2009). GPA and delinquency: Testing theories of delinquency through panel data. *Studies on Korean Youth*, *20*, 91–111.

Lubin, B., Larsen, R. M., Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist, 39,* 451-454.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults.* Baltimore, MD: Williams & Wilkins.

Marquez, K. (2015). Is violence necessary to the Korean education system? *10 Magazine.* Retrieved from http://www.10mag.com

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30,* 525-564.

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods, 2,* 248-260.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York, NY: Routledge.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial invariance on selection of multiple populations. *Psychological Methods, 9,* 93-115.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100,* 674–701.

Muthen, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes,* Version 5.

National Center for Juvenile Justice (2015). Juvenile Arrest Rates by Offense, Sex, and Race. Retrieved from http://www.ojjdp.gov/ojstatbb/crime/excel/JAR_2014.xls

Ng-Mak, D. S., Salzinger, S., Feldman, R. S., & Stueve, C. A. (2004). Pathologic adaptation to community violence among inner-city youth. *American Journal of Orthopsychiatry, 74,* 196–208.

Oh-Hwang, Y. (1993). Linkage between home environments, child psychosocial maturity, and child academic achievement. *Gifted International, 8,* 32-37.

Onifade, E., Petersen, J., Bynum, T., & Davidson, W. (2011). Multilevel recidivism prediction: Incorporating neighborhood socioeconomic ecology in juvenile justice risk assessment. *Criminal Justice and Behavior, 38,* 840–853.

Orpinas, P., Murray, N., & Kelder, S. (1999). Parental influences on students' aggressive behaviors and weapon carrying. *Health Education & Behavior, 26,* 774–787.

Quinsey, V.L., Skilling, T.A., Lalumiere, M. L., & Craig, W. M. (2004). Juvenile delinquency: Understanding the origins of individual differences. Washington, D.C: American Psychological Association.

Park S., & Kim, Y. (2016). Prevalence, correlates, and associated psychological problems of substance use in Korean adolescents. *BioMed Central, 16,* 79.

Park, M., Choi, J., & Lim, S. J. (2014). Factors affecting aggression in South Korean middle school students. *Asian Nursing Research, 8,* 247-253.

Rappaport, N., & Thomas, C. (2004). Recent research findings on aggressive and violent behavior in youth: Implications for clinical assessment and intervention. *Journal of Adolescent Health, 35,* 260–277.

Reagan, T. (1996). *Non-Western educational traditions: Alternative approaches to educational thought a proactive*. Mahwah, NJ: Lawrence Erlbaum Associates Publications.

Schmitt, N. (1996). Uses and abuses of Coefficient Alpha. *Psychological Assessment, 8*, 350-353.

Shin, S., & Koh, M. (2005). Korean education in cultural context. *Essays in education, 14*, 1-10.

Singelis, T. A., Triandis, H. C., Bhawuk, D. P. S., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism–collectivism: A theoretical and measurement refinement. *Cross-Cultural Research, 29*, 240–275.

Siu, S. F. (1992). *Toward an understanding of Chinese American educational achievement (Report No. 2).* Center on families, communities, schools, and children's learning.

Son, I. B., & Choi, J. A. (2013). Parental knowledge trajectory and its relationships with parental warmth and adolescent disclosure for Korean adolescents. *Children and Youth Services Review, 35*, 1643-1649.

Stone, G., Buehler, C., & Barber, B. K. (2002). Interparental conflict, parental psychological control, and youth problem behaviors. In B. K. Barber (Ed.), *Intrusive parenting: How psychological control affects children and adolescents* (pp. 53-96). Washington, D.C.: American Psychological Association

Supreme Prosecutors' Office of Korea. (2011). *Crime analysis* [Data file]. Retrieved from http://www.spo.go.kr/spo/info/stats/stats02.jsp

Tellegen, A., & Ben-Porath, Y. S. (1992). The new uniform T scores for the MMPI-2: Rationale, derivation, and appraisal. *Psychological Assessment, 4*, 145-155.

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

U.S. Department of Health & Human Services, Office of Adolescent Health (2015). *United States Adolescents Substance Abuse Facts.* Retrieved from http://www.hhs.gov

University of Minnesota Press (2016). MMPI-2-RF, MMPI-2, and MMPI-A translations. Retrieved July 30, 2016, from http://www.upress.umn.edu/test-division/translations-permissions/permissions.

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

Veltri, C. O., Graham, J. R., Sellbom, M., Ben-Porath, Y. S., Forbey, J. D., O'Connell, C., et al. (2009). Correlates of MMPI-A scales in acute psychiatric and forensic samples. *Journal of Personality Assessment, 91,* 288-300.

Wang, J. (2014). *Establishing measurement invariance of the MMPI-2 Restructured Clinical Scale 4 (Antisocial Behavior) using American and Korean clinical samples* (Unpublished master's thesis). Central Michigan University, Mount Pleasant.

Wang, P. W., Yang, P. C., Yeh, Y. C., Lin, H. C., Ki, C. H., Liu, T. L., et al. (2013). Self- esteem in adolescent aggression perpetrators, victims and perpetrator-victims, and the

moderating effects of depression and family support. *The Kaohsiung Journal of Medical Sciences, 29,* 221-228.

Webster, G. D., & Kirkpatrick, L. A. (2006). Behavioral and self-reported aggression as a function of domain-specific self-esteem. *Aggressive Behavior, 32,* 17-27.

Welsh, G. S. (1965). MMPI profiles and factors A and R. *Journal of Clinical Psychology, 21,* 43-47.

White, J. V. (2017) A comparative evaluation of Korean MMPI-A and MMPI-A-RF substance use scales. (Doctoral dissertation).

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10,* 395-426.

Won, H. J. (1989). *The daily leisure of Korean school adolescents and its relationship to subjective well-being and leisure functioning* (Doctoral dissertation, Department of Leisure Studies and Services, University of Oregon).

KRISTOFFER Y. PARK

2698 Claredon Trace
Kennesaw, GA 30144

Curriculum Vitae

kristoffer121@gmail.com
404.421.6085

EDUCATION

2013 – 2018 **Doctoral Candidate**
Virginia Consortium Program in Clinical Psychology, Norfolk, VA
Ph.D., Summer 2018
University-based, APA accredited program, jointly sponsored by:
Old Dominion University, Norfolk State University, Eastern Virginia Medical School
Dissertation: *Examining the measurement invariance of the Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF) Externalizing scales among Korean and American adolescent normative samples*
Faculty Advisor: Jennifer M. Flaherty, Ph.D.

2017 – 2018 **Aurora Community Mental Health Center, Pre-Doctoral Intern**
Child and Family Services, Asian Pacific Development Center, Intercept Day Treatment Program, Aurora, CO

Treatment Setting: Community Mental Health Center/Integrative Primary Care
Population: Children, adolescents, families, immigrants, and refugees, ages 5-75
Diagnoses: Mood Disorders, PTSD, Schizophrenia, Intellectual Disability, Autism Spectrum Disorder, Substance Use
Supervision: Danielle Mohr, Ph.D., Eri Asano, Ph.D., Margaret Charlton., Ph.D., 1 hour per week, 1 hour per week group supervision, 3 hours per week seminars/didactics
Hours: 40 hours on site per week

PEER-REVIEWED PUBLICATIONS

- Debb, S. M., Colson, D., Hacker, D., & **Park, K.Y.** (2018). Validating the Connor-Davidson Resilience Scale (CD-RISC) with third-year African American college students. *Journal of Negro Education.*

PRESENTATIONS

- **Park, K.Y.** (2015, September). *Acculturation and enculturation, perception of depressive symptoms, and help-seeking behaviors among Korean Americans.* Presented at the Annual Virginia Consortium Program in Clinical Psychology Research Day, Norfolk, VA.
- **Park, K.Y.** (2015, August). *Acculturation and enculturation, perception of depressive symptoms, and help-seeking behaviors among Korean Americans.* Poster presentation presented at the Asian American Psychological Association in Toronto, CA.
- **Park, K.Y.**, Parker, T. (2014, August). *The internal and external psychometric properties of the NEO Five-Factor Inventory-3 in a sample of college students at a historically black university.* Poster presentation presented at the American Psychological Association in Washington, D.C.