

**EQUIVALENCE AND FAKING ISSUES OF THE AGGRESSION
QUESTIONNAIRE AND THE CONDITIONAL REASONING TEST FOR
AGGRESSION IN KOREAN AND AMERICAN SAMPLES**

A Dissertation
Presented to
The Academic Faculty

by

Hye Joo Lee

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
May 2012

COPYRIGHT 2012 BY HYE JOO LEE

**EQUIVALENCE AND FAKING ISSUES OF THE AGGRESSION
QUESTIONNAIRE AND THE CONDITIONAL REASONING TEST FOR
AGGRESSION IN KOREAN AND AMERICAN SAMPLES**

Approved by:

Dr. Lawrence R. James, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Rustin Meyer
School of Psychology
Georgia Institute of Technology

Dr. Jae Yoon Chang
Department of Psychology
Sungshin Women's University

Dr. Charles Parsons
College of Management
Georgia Institute of Technology

Dr. Jack Feldman
School of Psychology
Georgia Institute of Technology

Date Approved: January 27, 2012

ACKNOWLEDGEMENTS

I praise the Lord and glorify Him for all of His guidance in the pursuit of all of my accomplishments.

I cannot express my gratitude enough to my advisor, Dr. Lawrence James, for his support, help, caring, patience, and valuable advice. I would not have been able to finish the Ph.D. program without the guidance of him. I would also like to thank Dr. Chang, who has very graciously helped me collect data, served on my committee, and provided me with unwavering support and guidance. I could not have completed my dissertation without his invaluable assistance. I would also like to thank Dr. Feldman, Dr. Meyer, and Dr. Parsons for their comments, suggestions, and service on my committee.

I am extremely grateful to my parents and sister (Evelyn) for their sacrifices, support, love, and trust. They never wavered in their encouragement of me throughout this long process. I would also like to extend my thanks to Pastor Choi and Sugarloaf Korean Baptist Church community members, who guided me spiritually and prayed for me whenever I faced my greatest challenges and frustrations.

“But he knows the way that I take; when he has tested me, I will come forth as gold.”

(Job 23:10)

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	viii
 <u>CHAPTER</u>	
1 CHAPTER 1 INTRODUCTION	1
2 CHAPTER 2 MEASURES OF AGGRESSION	3
Self-Attributed Aggression	4
Implicit Personality Assessment of Aggression	6
3 CHAPTER 3 EQUIVALENCE OF THE AGGRESSION QUESTIONNAIRE	10
Construct Bias	11
Method Bias	12
Item Bias	15
4 CHAPTER 4 EQUIVALENCE OF THE CRT-A	17
Construct Bias	17
Method Bias	20
Item Bias	21
More Measurement Issues Concerning the Korean CRT-A	21
5 CHAPTER 5 METHODOLOGY	24
Study 1 (Equivalence of the AQ and the CRT-A)	24
Participants	24
Procedure	24

Data analysis	27
6 CHAPTER 6 RESULTS AND DISCUSSION	30
Results	30
Mean Number of Respondents Using the Midpoint and Extreme Points (1 or 7)	30
Factor Analyses of the AQ	31
DIF Analysis of the AQ	36
Factor Analyses of the CRT-A	43
DIF Analyses of the CRT-A	47
Discussion	54
7 CHAPTER 7 METHODOLOGY	57
Study 2 (Test of Faking)	57
Participants	57
Procedure	58
Data analysis	58
8 CHAPTER 8 RESULTS AND DISCUSSION	60
Results	60
Discussion	61
9 CHAPTER 9 GENERAL DISCUSSION	63
Contributions and Implications	65
Limitations	67
Future Directions	68
REFERENCES	70

LIST OF TABLES

	Page
Table 1: Justification Mechanisms for Aggression	8
Table 2: Illustrative Conditional Reasoning Problems	26
Table 3: Descriptive Statistics for the Aggression Questionnaire in Study 1	31
Table 4: Results of Confirmatory Factor Analysis for the Korean Aggression Questionnaire	32
Table 5: Exploratory Factor Analysis for English Aggression Questionnaire. Factor Loadings	34
Table 6: Exploratory Factor Analysis for Korean Aggression Questionnaire. Factor Loadings	35
Table 7: IRT adjusted threshold parameters of the Aggression Questionnaire items between Korean and American	38
Table 8: Exploratory Factor Analysis for English Conditional Reasoning Test-Aggression. Factor Loadings	45
Table 9: Exploratory Factor Analysis for Korean Conditional Reasoning Test-Aggression. Factor Loadings	46
Table 10: IRT adjusted threshold parameters of the Conditional Reasoning Test-Aggression items between Korean and American	49
Table 11: Descriptive Statistics for the Conditional Reasoning Test-Aggression in Study 2	61

LIST OF FIGURES

	Page
Figure 1: Item Characteristic Curves for Aggression Questionnaire Item 9: Korean versus American	39
Figure 2: Item Characteristic Curves for Aggression Questionnaire Item 17: Korean versus American	40
Figure 3: Item Characteristic Curves for Aggression Questionnaire Item 19: Korean versus American	41
Figure 4: Item Characteristic Curves for Aggression Questionnaire Item 24: Korean versus American	42
Figure 5: Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 24: Korean versus American	50
Figure 6: Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 16: Korean versus American	51
Figure 7: Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 18: Korean versus American	52
Figure 8: Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 25: Korean versus American	53

SUMMARY

Researchers have raised concerns about measurement equivalence in comparing personalities across cultures using personality assessments. The self-reported personality measurements often do not assess the same construct, trigger different response styles (i.e., extreme response style), or use behavioral exemplars that are inappropriate across cultures (Byrne & Watkins, 2003; Chen, 2008; Poortinga, van de Vijber, & van Hermert, 2002, van de Vijver & Leung, 1997). James et al. (2005) developed a new measurement system for aggression that is different from traditional personality assessment. It is referred to as the Conditional Reasoning Test for Aggression (CRT-A). The CRT-A is an indirect measure for assessing unconscious motives to be aggressive that was developed in the USA. It has not been studied with people from different cultures. Study 1 investigated the equivalences of the Aggression Questionnaire (AQ) and the CRT-A by administering both to groups of Americans (n=432) and Koreans (n=363). Results based on the exploratory and confirmatory factor analyses and DIF analyses showed that the AQ and CRT-A are not invariant across these cultures. Study 2 replicated LeBreton et al.'s (2007) study regarding faking issues of the CRT-A with the Korean population. Study 2 found that on the CRT-A, Koreans were able to identify aggressive alternatives when they were told to do so, and Korean students and employees did not score differently on the CRT-A. Implications and future directions of the study are discussed herein.

CHAPTER 1

INTRODUCTION

Individuals are likely to understand themselves through comparison others (Festinger, 1954). For example, students understand how well they are doing in class by comparing their scores with class average scores, and politicians are likely to understand where their country stands economically by comparing their gross domestic product (GDP) with that of other countries. Similarly, cross-cultural psychologists understand how the personality traits of various cultures differ by comparing their scores on personality questionnaires such as the NEO-PI-R (McCrae, 2002). Nevertheless, unlike those comparing test scores or GDPs across cultures, those comparing personality traits might not be able to make the same inferences from scores on personality assessments because of variances in constructs, methods, and instruments. For instance, if Koreans score lower on the Aggression Questionnaire (AQ) than Americans do, the scores do not necessarily indicate that Koreans are less aggressive than Americans are. Their lower scores on the AQ could be the result of their tendency to choose midpoints compared to Americans' tendency to choose extreme points on questionnaire items. Therefore, without a test of measurement invariance, inferences made from the group-level comparison could be meaningless (e.g., Byrne & Watkins, 2003; Chen, 2008; van de Vijver & Leung, 1997).

This study conducts measurement variance (i.e., differences between the responses of Americans and those of Koreans) on explicit aggression measures, which

are mostly based on rating scales, and an implicit aggression measure, which has a multiple-choice format. This study addresses the following questions:

- 1) Are the Korean and English versions of the AQ (i.e., a self-report measure of aggression) equivalent?
- 2) Are the Korean and English versions of the Conditional Reasoning Test for Aggression (CRT-A; i.e., implicit measure of aggression) equivalent?

The study also addresses a subsequent question regarding the CRT-A:

- 3) Can Koreans fake their responses on the CRT-A?

To answer the above questions, Chapter 2 examines widely-used aggression measures in the United States, Chapter 3 reviews three sources of bias—construct, method, and item—in the AQ, and Chapter 4 reviews the same sources of bias with the CRT-A in addition to faking issues associated with the CRT-A. Next, Chapter 5 discusses the methodology of Study 1, and Chapter 6 presents results and discussion of Study 1. Subsequently, Chapter 7 and Chapter 8 describe the methodology of Study 2 and its results and discussion, respectively. Finally Chapter 9 provides a general discussion of this study with implications, limitations, and future directions.

CHAPTER 2

MEASURES OF AGGRESSION

“Aggressiveness evolves from a desire or motive to overcome opposition forcefully, to fight, to revenge an injury, to attack another with intent to injure or kill, and to oppose forcefully or punish another (Murray, 1938)” (James & Mazerolle, 2001, p. 8). One of the most undesirable characteristics, aggressiveness is strongly associated with anything from minor unwanted behaviors such as lying (Bing, Stewart, Davison, Green, McIntyre, & James, 2007; Russell & James, 2008), sabotage (James, McIntyre, Glisson, Bowler, & Mitchell, 2004), absenteeism (James et al., 2005; Hogan & Hogan, 1989; Patton, 1999), grievances (Hogan & Hogan, 1989), cheating (Russell & James, 2008), and traffic violations (Bing et al., 2007), to serious evil behaviors such as stealing (Sablinski & Mitchell 2006), fighting (Gentile, Lynch, Linder, & Walsh 2004), and physical attacks (Frost, Ko, & James, 2007). Thus, the personality construct of aggression and its assessments has been the focus of considerable interest.

A number of methodologies such as behavioral measures and observation techniques have been proposed to study aggression. However, the approach that has attracted the greatest interest is the self-report method such as the questionnaire (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007), which has been used to study neuroticism, the most closely related subset of the Big Five traits to aggression (e.g. Bing et al., 2007; Frost, Ko, & James, 2007; James et al., 2005). While child aggression is often assessed by teacher and peer evaluations (e.g. Dodge & Frame, 1982; Huesmann & Guerra, 1997), adult aggression usually relies on self-report

assessments. However, self-reporting may not produce an accurate assessment of an individual's aggressiveness not only because individuals may not be able to perceive their own aggressive tendencies but also because they are generally less likely to report their aggressiveness in stressful situations (e.g., job applications; Rosse, Stecher, Miller, & Levin, 1998). Thus, self-report methods tend to generate information about how one perceives his/her own aggression or how one wants to be perceived rather than a true representation of one's true aggressive disposition. Furthermore, self-reported aggression may not capture multiple facets of aggression. For example, self-attributed aggression more likely taps the explicit level of aggression, not the unconscious level of aggression. However, unlike explicit aggression, implicit aggression measures seem to access the unconscious level of aggression. This paper begins with a discussion of three popular self-report measures of aggression and a new indirect assessment of aggression: the Conditional Reasoning Test for Aggression (CRT-A).

Self-Attributed Aggression

NEO Personality Inventory-Revised (NEO-PI-R)

One of the most popular self-report measurements of the Big Five personality traits is the NEO-PI-R (Costa & McCrae, 1992). The NEO-PI-R consists of 243 subjective items assessing an individual's level of extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. A short version of the NEO-PI-R is the NEO Five-Factor Inventory (NEO-FFI, McCrae & Costa, 2004), which reduces the number of items to 60. Respondents indicate to what extent they agree with each statement based on a 5-point Likert scale. Self-attributed aggression is assessed by the neuroticism scale of the NEO-PI-R. Underlying facets of neuroticism are anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Sample

items of neuroticism are “I get upset easily,” “I worry about things,” and “I have frequent mood swings.” The neuroticism scales of the NEO and the NEO-FFI have shown internal consistent reliability of 0.92 and 0.79-0.85, respectively. The test-retest reliability of the NEO-neuroticism was 0.87 over a three-month interval and 0.82 with NEO PI-R neuroticism over a six-year interval. Also, the NEO-PI-R demonstrated good convergent and discriminant validity. Neuroticism is negatively related to job satisfaction (Judge, Heller, & Mount, 2002), academic satisfaction (Trapmann, Hell, Hirn, & Schuler, 2007), performance motivation (Judge & Ilies, 2002), and positively related to interpersonal and organizational deviance (Berry, Ones, & Sackett, 2007).

The Aggression Questionnaire (AQ)

The Aggression Questionnaire (AQ) was developed by Buss and Perry (1992) and contains 29 items that assess four facets of aggression: physical aggression, verbal aggression, anger, and hostility. Sample items of these four facets are “Once in a while I can't control the urge to strike another person,” “I can't help getting into arguments when people disagree with me,” “Some of my friends think I'm a hothead,” and “I sometimes feel that people are laughing at me behind my back,” respectively. Internal consistency of each subset ranged from 0.72 to 0.85, and the test-retest reliability over a nine-month interval ranged from 0.72 to 0.80. The discriminant validity of sociability ranged from -0.12 to 0. Scores on the AQ were significant associated with dominance, sexual jealousy (Archer & Webb, 2006), a Type-A behavior pattern (Innamorati, Pompili, Ferrari, Cavedon, et al., 2006), bullying (Ireland & Archer, 2004), and eating disorders (Miotto, De Coppi, Frezza, Petretto, Masala, & Preti, 2003).

Jackson Personality Research Form (PRF)

Jackson's (1984) Personality Research Form (PRF), which has been used for over forty years, is a still a valid and highly cited measure. The PRF comes in various forms,

the most popular version is E, which consists of 352 items assessing 22 dimensions of personality-abasement, achievement, affiliation, aggression, autonomy, change, cognitive structure, dependence, desirability, dominance, endurance, exhibition, harm avoidance, impulsivity, infrequency, nurturance, order, play, sentience, social recognition, succorance, and understanding. Unlike the aforementioned personality questionnaire, the items are based on true/false responses. The PRF has acceptable internal consistency ranging from 0.50 to 0.91: test-retest reliabilities over a two-week interval range from 0.80 to 0.96. An aggression scale from the PRF has shown successful validities with self-reported likelihood to rape (Greendlinger & Byrne, 1987), group task accomplishment (Brenner & Tomkiewicz, 1980), cheating behavior (Kelly & Worell, 1978), paranormal beliefs (Auton, Pope, & Seeger, 2003), and an exploiting/active lifestyle (Wheeler, & Acheson, 1993).

Implicit Personality Assessment of Aggression

The Conditional Reasoning Test for Aggression (CRT-A)

Individuals routinely perform activities based on what they believe is right or appropriate. This judgment, belief, or idea is not the same for everybody. Even in the same situation, people can make different judgments, and they act accordingly. Even if the actions or judgments may not seem acceptable or reasonable to others, most individuals are ready to justify their actions. Thus, aggressive individuals and non-aggressive individuals make different decisions in similar situations, and both parties have reasons for their actions that seem reasonable and rational to them. The reasoning biases that aggressive individuals use to make their actions appear rational and sensible are called “Justification Mechanisms” (JMs; James 1998). James classified implicitly aggressive individuals’ biases into six JMs (Table 1): hostile attribution bias, potency bias, retribution bias, victimization by powerful others bias, derogation of target bias, and social discounting bias. More details about each JM will be discussed in Chapter 4. JMs

are based on theory from previous research, not on empirical results exclusively. James contends that JMs are in place for implicitly aggressive individuals' reasoning processes. These individuals are not only aggressive but also ready to justify their aggressive dispositions. These processes tend to happen outside of their awareness. Based on the six JMs, the CRT for Aggression (CRT-A) consists of what appears to be 22 inductive reasoning items, with three bogus items included for face validity. Each item has a short premise followed by four alternatives: One alternative is attractive to implicitly aggressive individuals, one is a pro-social alternative, and two are illogical alternatives. Individuals who endorse an aggressive response will score +1, a pro-social response will score -1, and an illogical response will score 0. James and his colleagues validated the measure, which showed promising validity in predicting employee absenteeism; counterproductive behaviors such as a theft, sabotage, and work performance (James et al., 2004); perception of injustice (Burroughs, 2001); and obstructionism by basketball players (Frost, Ko, & James, 2007).

Table 1. Justification Mechanisms for Aggression

1. Hostile attribution bias's core is an implicit assumption that (like oneself) people tend to be motivated by a desire to harm others (Anderson, 1994; Tedeschi & Nesler, 1993; Toch, 1993). This latent bias is instrumental in shaping conscious attempts to explain why others behave as they do. Such explanations show a strong predilection to attribute behavior to malevolent purpose and harmful intent (cf. Crick & Dodge & Coie, 1987). Even benign or friendly acts may be credited to hidden, hostile agendas designed to inflict harm. The attributions of hostile intent are central to the aggressive person's attempts to rationalize his or her own hostile behaviors as acts of self-defense intended to ward off physical or verbal attack.
2. Potency bias is grounded in the implicit assumption that interactions with others are contests to establish dominance versus submissiveness (Anderson, 1994; Gay, 1993; Millon, 1990). This bias unconsciously shapes framing; the actions of others pass through a perceptual prism primed to distinguish (a) strength, assertiveness, dominance, daring, fearlessness, and bravery from (b) weakness, impotence, submissiveness, timidity, compliance, and cowardice (James & Mazerolle, 2002). Such framing promotes reasoning that the use of aggression to dominate others demonstrates strength, bravery, control, and fearlessness. Not active person may thus rationalize aggression by reasoning (a) that aggression in an act of strength or bravery that gains respect from others and (b) that to show weakness is to invite powerful others to take advantage of you.
3. Retribution bias centers on an implicit assumption that exacting retribution is of greater consequence than preserving or maintain a relationship. This bias surfaces as a proclivity to favor retaliation as a more rational behavior than reconciliation (cf. Bradbury & Fincham, 1990; Dodge, 1986; Laursen & Collins, 1994). For example, aggression is seen as justifiable if it is intended to restore or to exact retribution for a perceived wrong. Retaliation is thus assumed to be more reasonable than forgiveness, vindication appears more reasonable than reconciliation, and obtaining revenge appears more reasonable than maintaining a relationship. This bias often underlies justifications for aggression engendered by wounded pride, challenged self-esteem, and perceived disrespect (cf. Baumeister, Smart, & Boden, 1996).
4. Victimization by powerful others bias has an a nucleus an implicit assumption that the powerful will inflict harm of the less powerful (Averill, 1993; Finnegan, 1997; Toch, 1993). This assumption underlies a conscious proclivity to see oneself as the victim of inequity, exploitation, injustice, and oppression by those who are more powerful in one's life (e.g., parents, teachers, supervisors, employing organizations, or institutions such as the Internal Revenue Service). Framing of events, hypotheses about cause and effect, and confirmatory searches for evidence both engender and reinforce inferences that people are being victimize by powerful others. This reasoning furnishes the foundation for justifying acts of aggression as warranted corrections of inequities or legitimate strikes against oppression.

5. Derogation of target bias consists of an unconscious tendency to characterize those one wishes to make (of has made) targets of aggression as evil, immoral, or untrustworthy (cf. Wright & Mischel, 1987). To infer or associate such traits with a target makes the target more deserving of aggression.
 6. Social discounting bias has at heart an implicit assumption that social customs restrict free will and the opportunity to satisfy needs. Reasoning shaped by this latent bias reflects disdain for traditional ideals and conventional beliefs (cf. Finnegan, 1997; Loeber & Stouthamer-Loeber, 1998; Millon, 1990). For example, attempts to identify the most logically plausible causes of social events typically lean toward the cynical and critical. Reasoning will further evidence a lack of sensitivity, empathy, and concern for social customs, often accompanied by the absence of rational prohibitions against behaving in socially unorthodox ways. Socially deviant behavior intended to harm others is rationalized by inferring that it allows one to attain freedom of expression, release from the shackles of social customs, and liberation from confining social relationship.
-

Sources: James, R. L., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., Frost, B. C., Russell, S. M., Mitchell, T. R. & Williams, L. J. (2005). A Conditional Reasoning Measure for Aggression. *Organizational Research Methods*, 8, 69-99

CHAPTER 3

EQUIVALENCE OF THE AGGRESSION QUESTIONNAIRE

The self-reported aggression measures (the AQ and the NEO-PI-R) reviewed in Chapter 2, which have been translated into several languages, demonstrated promising validity with populations including the Chinese subjects (Ang, 2007) as well as Hungarian (Gerevich, Bácskai, & Czobor, 2007), British (Archer, Holloway, & McLoughlin, 1995), Italian (Fossati, Maffei, Acquarini, & DiCeglie, 2003), and Japanese (Nakano, 2001) respondents. The NEO-PI-R has also been validated with Turkish subjects (Gülgöz, 2002), as well as French (McCrae, Costa, Pilar, Rolland, & Parker, 1998), Indian (Lodhi, Deo, & Belhekar, 2002), Korean (Piedmont & Chae, 1997), and Zimbabwean subjects (Piedmont, Bain, McCrae, & Costa, 2002). Furthermore, the cross-cultural studies also supported that collectivist cultures tend to show lower levels of aggression than individualistic cultures (Bergeron & Schneider, 2005). For instance, Israeli Jews, known to be low collectivists, showed lower levels of indirect aggression and higher levels of direct aggression than Israeli Arabs, known to be high collectivists, based on the Workplace Aggression Tolerance Questionnaire (WATQ; Galin & Avraham, 2009).

Group level comparisons of aggression levels provide meaningful information in terms of where each culture stands compared to other cultures. However, without a thorough investigation of measurement equivalence, particularly measurements that use rating scales, the comparison may lead to erroneous conclusions because of a number of sources of biases. Thus, to confirm the conclusion that collectivist cultures demonstrate lower levels of aggression, the WATQ should have tested for measurement equivalence across cultures. In this chapter, I will review the possible sources of bias on the AQ, the most widely used measure to assess aggression when the questionnaire is applied to

Americans (individualists) and Koreans (collectivists). Researchers appear to agree on three sources of bias that contribute to measurement variance: 1) construct bias, 2) methodology bias, and 3) item bias (Byrne & Watkins, 2003; Church, 2001; van de Vijver & Leung, 2001).

Construct Bias

Construct bias occurs when the sampling behaviors of the aggression construct between cultures do not overlap. The two factors that contribute to the construct bias suggested by Byrne and Watkins (2003) are “First, the behaviors being tapped as indicators of a construct can be differentially appropriate across cultural groups...Second, the extent to which all relevant dimensions of the construct have been included in the formulation of item content varies across groups” (p.157). To be equivalent measures of aggression, the measures should have similar identifiers capturing the aggression construct and the same number of underlying dimensions.

From the three measures of aggression, the NEO PI-R is the only measure that has been tested in a Korean population. Piedmont and Chae (1997) conducted a cross-cultural study in which 654 Koreans took the Korean version of the NEO PI-R. This group showed acceptable reliability and validity. In the second study, 116 bilingual Koreans took both English and Korean versions of the NEO PI-R, and their responses on both measures were comparable. The cross-cultural study of the NEO-PI-R confirmed that the construct of personality is appropriate to Korean samples and presented a clear five-factor structure, the same number of underlying dimensions as for Americans. Currently, the Korean version of the NEO PI-R is widely used in assessing the Big Five Personality traits and in investigating their relationship with Korean employees' task performance (Lee^a, 2001), coping styles (Roesch, Wee, & Vaughn, 2006), and English proficiency (Lee^b, 2001).

The AQ, another measure of aggression, with 29 items, captures four underlying factors: physical and verbal aggression, anger, and hostility (Buss & Perry, 1992); and a short version of the AQ with 12 items also confirmed the four-factor structure (Bryant & Smith, 2001). The equivalency of the AQ has not been tested on Koreans; therefore, in this study, I will focus on the AQ as an explicit measure for aggression. Although the AQ has not been validated within the Korean population, the AQ has shown promising validity and reliability when administered to Asian populations (i.e., Chinese and Japanese). For example, 967 Chinese male prisoners showed comparable four-factor structures on the short form of the AQ, and their scores on the AQ were higher than the normal group, as expected. Furthermore, after the exclusion of two items, the Japanese population also demonstrated the psychometric properties of the Japanese version of the AQ (Nakano, 2001). Koreans have assessed aggression using the Korean Aggression Questionnaire, which shows similar factor structures, including physical, verbal, and indirect aggression, hostility, and anger.

Method Bias

Church (2001) introduced three types of method bias: 1) sample bias, 2) instrument bias, and 3) administration bias. The various education levels of the sample respondents and different study procedures with samples from diverse cultures led to sample bias. However, sample bias can easily be controlled by matching the education level of participants, such as comparing college students from two different countries. One can follow exactly the same instructions to remove the administration bias. A more problematic bias is instrument bias, which can lead to varying response styles (e.g., an extreme response style), especially when one uses Likert scale (Poortinga, van de Vijber, & van Hemert, 2002).

According to Clark (2000), the “extreme response style was identified by Cronbach (1946) as the tendency for some individuals to consistently use the extreme

ends of response scales in a multiple response category format” (p. 138). This extreme response style is prevalent in cross-cultural studies; those from certain cultures tend to use the extreme end of categories, while those from other cultures are likely to choose mid-point scales more consistently. Chen, Lee, and Stevenson (1995) explored the different styles of responses on rating scales of four different countries—Japan, China, Canada, and the United States—and found that the response styles of these cultures significantly differed. For example, students from the two Eastern countries demonstrated a higher mean score of use of midpoint scale values on the items of orientation toward individualism and collectivism than students from the two North American countries. Furthermore, American students used significantly more extreme values than Japanese, Chinese, and Canadian students. The different response styles could be due to cultural differences such as modesty, typical in Asians but not in other groups (Chen, Lee, & Stevenson, 1995).

Another possible source of method bias that creates difficulty for cross-cultural researchers making a comparison or inferences is the reference group effect (RGE: Heine, Lehman, Peng, & Greenholtz, 2002; Peng, Nisbett, & Wong, 1997), which is described as “the tendency for people to respond to subjective self-report items by comparing themselves with implicit standards from their culture” (Heine et al., 2002; Heine, Buchtel, & Norenzaya, 2008). As the reference group effect is a relatively new finding, it has not been considered a source of bias. However, in this author’s opinion, the reference group effect must be discussed within the context of any cross-cultural study. The RGE is based on Festinger’s (1954) theory of social comparison, which declared that people tend to understand themselves by comparing themselves to others. For example, students in a class understand how they are performing by comparing their scores with the class average or with the student sitting next to them. Heine, Buchtel, and Norenzayan (2008) provided evidence for implicit comparison using a subjective Likert scale on a domain of conscientiousness with samples from a variety of countries. They argued that since

people unconsciously compare themselves with others around them, self-reported personality measures are less valid in predicting objective criteria when cross-culturally compared. For instance, occupational success has been shown to be correlated with conscientiousness (Judge, Higgins, Thoresen, & Barrick, 1999), and Heine et al. used the GDP as an index of occupational success for each country and assessed conscientiousness using the NEO-PI-R (McCrae, 2002) with samples from 17 to 55 countries. They found a significant negative correlation ($r = -.66$) between aggregated conscientiousness scores and GDP and concluded that individuals from various countries responded to the NEO-PI-R using their own reference group or a reference group from their countries; therefore, the aggregated scores of conscientiousness were not correlated with GDP, as expected. In other words, the country with the highest GDP did not have the highest score on conscientiousness. Thus, even though the people in the country with the highest GDP were expected to be highly conscientious, because they compared themselves with others from their own country, the aggregated conscientiousness score was not significantly higher than that of people from other countries. Heine et al. clearly demonstrated that even individuals from different countries unconsciously compared themselves with others when they respond to the subjective Likert scale.

Crede, Bashshur, and Niehorster (2010) recently claimed that explicit instructions using a specific reference group could change respondents' scores on a self-reported personality measure. Their argument states that the choice of reference groups is usually unconscious and implicit but that respondents' choice of reference groups can be cued by instruction and that scores provide meaningfully different information depending on the reference group. For instance, Crede, Bashshur, and Niehorster (2010) asked participants to complete a 10-item measure of conscientiousness from the International Item Pool (Goldberg et al., 2006) using four reference groups: 1) their immediate families, 2) people of the same age and gender, 3) close friends and peers, and 4) people in general. Respondents were also asked to rate themselves without any and also without specified

comparison group (reference-free). The results provided significant mean score differences between reference groups. The respondents' scores were lowest when they were compared with their immediate family and highest when they were compared with no reference group or when they were compared with people of the same age and gender. Interestingly, the scores with no reference group and a reference group in general significantly differed. The study by Crede, Bashshur, and Niehorster (2010) illustrated that people implicitly make comparisons as they respond to subjective personality surveys, and that they can be cued by instruction to compare themselves with different groups of people.

Item Bias

"Item bias" refers to item-level misrepresentation that could be the result of a poor translation process or irrelevant behavioral samples (Byrne & Watkins, 2003). The very first step of the cross-cultural study is translation. If one does not translate the instrument thoroughly, then the instrument will not assess what it is supposed to assess. The most widely used translation technique in cross-cultural studies is back-translation (i.e., Aycicegi, Dinn, & Harris, 2005; McCrae & Costa, 1997, Noh, Avison, & Kaspar, 1992, Schmitt & Allik, 2005). However, Barger, Nabi, and Hong (2010) argued that the back-translation procedure does not adequately capture the concept of emotion. For example, if the word for *disgust* has not been accurately translated in a Chinese questionnaire, the inaccurate translation can warp the results of a cross-cultural study. The AQ contains several items that describe emotion such as "I let my anger show when I do not get what I want " and "I wonder why sometimes I feel so bitter about things." Thus, a translator must pay extra attention when finding a word in one language that accurately reflects an emotional concept of another language. If researchers adapt additional techniques such as bilingual testing and retesting, proofreading of a translated language, or selecting the

best items after multiple people translate the measure they will provide a more accurate translation overall.

Even after an accurate translation procedure, samples of a behavior described in the measure may not be applicable to another culture, and inappropriate samples of a behavior can create an item bias. The sample behaviors on the AQ seem comparable to those listed on the KAQ. For example, "I have become so mad that I have broken things," "I like to play practical jokes," and "At times I get very angry for no good reason," are comparable with "When I am very mad, I slam doors," "I feel better after I play practical jokes," and "I often get angry about very small things," respectively.

Considering the biases mentioned above, this study proposes two research questions:

Question 1: Is the Korean AQ equivalent to the English AQ?

Question 2: Do Koreans use more of the midpoints of scale than Americans?

CHAPTER 4

EQUIVALENCE OF THE CRT-A

Although the development of equivalent measures across all cultures would be ideal, it is not practical, and most of the measures currently used contain several flaws in terms of measurement invariance. To solve the problem of non-equivalent measures, researchers have suggested several methods, one of which is to eliminate items that are invariant (Cheung & Rensvold, 2000; Dumka, Stoerzinger, Jackson, & Roosa, 1996). In addition, suggestions for reducing method bias due to social desirability are to use items that are equally socially desirable (Nederhof, 1985) or to use items with forced-choice options (Aupperle, 1984). The strictest solution is not to make comparisons (Chen, 2008; Cheung & Rensvold, 2000). However, before researchers become too pessimistic about cross-cultural comparison, the author wishes to explore the measurement equivalence of the CRT-A. Since the CRT-A is not based on rating scales or self-reports, it may not create the same problems as other self-report measures. Thus, this chapter will evaluate the same three sources of bias discussed in the previous chapter using the CRT-A.

Construct Bias

As mentioned, the use of biases or JMs (justification mechanisms) provides reasoning that sounds neither logical nor reasonable to pro-social individuals; however, to aggressive individuals, reasoning based on JMs appears to be sensible and rational. More interestingly, because these biases are implicit, aggressive individuals can use them to unconsciously rationalize their actions and beliefs. Examples of the use of JMs are presented below.

One example of an implicit bias of aggressive individuals who are often not aware of their aggressiveness is the latent hostile attribution bias (Dodge & Coie, 1987).

These individuals assume that all people have an innate motivation to harm others (Anderson, 1994; Toch, 1993), so they may view others' kind and polite gestures as hostile or malevolent. Furthermore, the illogical biases of implicitly aggressive individual influence their view of the well-mannered behaviors of others. Therefore, they think others hide their hostility and mask their harmful intent. Aggressive individuals' belief in being hostile to others is appropriate to them because even though others seem to be nice and kind, their true intent is perceived as malevolent and harmful. Similar to the latent hostile attribution bias, the feeling of being victimized by others in power is common in implicitly aggressive individuals. In other words, they believe that those in a powerful position take advantage of subordinates, who then become victims of the powerful people. While aggressive individuals believe they are simply victims of supervisors, teachers, parents, or others in positions of power, individuals who do not have such an implicit bias will view them as mentors or simply people trying to help or advise them. As a result, instead of respecting and obeying people in a higher position, aggressive individuals will regard them as unfair, justifying aggression towards them as a protective action.

A similar bias that aggressive individuals may hold is the derogation of target bias. They tend to see individuals they would like to be aggressive towards, or a target of aggression, as evil or immoral. Even if a target means no harm or wishes no malice toward the aggressive individual, the latter believes the target is untrustworthy without any logical or sensible reason. Therefore, any act of aggression toward the target is rational because the target is immoral or evil.

Another implicit bias of aggressive individuals is the potency bias. Implicitly, aggressive individuals think that relationships with others are a form of competition in which one is either dominant or submissive (Anderson, 1994; Gay, 1993). They believe that if they establish a friendly relationship with others, they become submissive to others and no longer maintain a position of control over them. That is, they are more likely to

consider a relationship as hierarchical rather than parallel, and as such, they prefer to be at the top of the hierarchy because the alternative is to be at the bottom. Therefore, they justify their aggression by claiming that if they are not forceful or belligerent, they will be considered cowardly or timid. A similar bias of aggressive individuals is the retribution bias, a latent bias that dictates that unconsciously aggressive people believe an appropriate or logical way to resolve issues with others is through retribution, not reconciliation (James et al., 2005). Although non-aggressive individuals try to maintain relationships with others through forgiveness and reconciliation, aggressive individuals believe others who hurt or frustrate them need to experience the same feelings as they felt. Therefore, believing revenge or retribution are logical and rational ways to resolve conflicts with others and maintain relationships, aggressive individuals do not see reconciliation or forgiveness as an option for maintaining a relationship with others.

The last bias, unlike the others, is the social discounting bias. Aggressive individuals believe that social rules or customs interfere with their freedom to express their ideas and their social needs. Instead of viewing social policies as necessary for maintaining peace in their communities, aggressive individuals think they restrict their free will. Such individuals believe that if they are to experience unrestricted freedom, they must act aggressively by going against the rules.

The six JMs make up the core principles of the CRT-A, but very few studies on implicit or passive aggression have been conducted with Koreans. However, the six JMs of the CRT-A seem to be valid for Koreans. According to Woo (2009), Koreans with passive aggressive personality disorders tend to complain a lot, have a discounting bias towards people in higher positions of authority, and tend to be negative and argumentative, withhold information, and sabotage relationships. These behaviors were also characteristic of implicitly aggressive individuals in the United States. Thus, the CRT-A is expected to assess the same construct, that of implicit aggression, with Koreans as well.

In terms of the factor structure for the CRT-A, a principal component analysis with Promax rotation showed that the CRT-A assesses three types of implicit aggression: external controls, internal controls, and powerlessness (Ko, Thompson, Shim, Roberts & McIntyre, 2009). The JMs that fall under external controls are victimization by powerful others and exploitation by societal norms; those that fall under internal controls are potency, dominance, and retribution. Powerlessness is described as a “lack of influence” (Ko et al., 2009). Eleven items represent the external controls, six items represent internal controls, and five items represent helplessness. Because internal controls and powerlessness do not contain enough items that explain these factors, this study developed new items that fall under these two factors for the Korean test.

Method Bias

The CRT-A consists of 22 inductive items in a format widely used across cultures for reasoning problems. One answer is an implicitly aggressive alternative, another answer is a pro-social alternative, and two of the alternatives are illogical. As no cross-cultural effect of multiple choice tests (e.g., the SAT, the GRE, the GMAT) taken by U.S. university-bound students have been found, the author does not expect cultural effects to play a role in the results of the CRT-A multiple choice questionnaires that will be administered to the Korean participants. However, the illogical choice on the CRT-A may cause cultural problems. Although the illogical alternatives clearly appear to be nonsensical to American students with a fifth grade or higher reading level (James & McIntyre, 2000), these same responses may not be illogical to Koreans. In one cross-cultural study of CRT-Relative Motive Strength (CRT-RMS, which has the same concept as the CRT-A in assessing motives to achieve) with 188 Korean students, more than half of Korean college students believed that the illogical answers were the most logical answers on two items of the CRT-RMS (Lee, 2009).

Item Bias

The last source of bias, item bias, can occur on the item level, for several of the items may not capture Koreans' implicit aggressiveness. Unlike the self-reported aggression questionnaire, which lists only sample aggressive behaviors, each of the CRT-A items starts with a short premise and offers four alternatives. Compared to the AQ, each of the premises in the CRT-A contains fewer emotional words, but premises from two CRT-A problems may not be appropriate to Koreans. One of the premises starts with "More people are getting permits to carry guns." This premise does not seem to be applicable to Koreans because no one is allowed to carry a gun nor obtain a permit to carry on in Korea. Another premise states that "American cars have gotten better in the last 15 years. American car-makers started to build better cars when they began to lose business to the Japanese. Many American buyers thought that foreign cars were better made." The AG alternative is "American car makers built cars to wear out 15 years ago, so they could make a lot of money selling parts" while the pro-social alternative is "The Japanese knew more than Americans about building good cars 15 years ago." This item is written from an American's perspective and contains a cultural issue; therefore, an aggressive or pro-social motive may not work in the same way as it would for Koreans. Furthermore, the attitudes of Koreans towards Japanese people differ from those of Americans (i.e., many Koreans' feel hostile towards the Japanese as a result of historical events); therefore, this item may assess cultural views towards the Japanese rather than implicit aggression.

The biases discussed above raise the following research question:

Question 3: Is the Korean CRT-A equivalent to the English CRT-A?

More Measurement Issues Concerning the Korean CRT-A

One of the strengths of the CRT-A is that it does not allow faking or responses that are simply socially acceptable (LeBreton, Barksdale, Robin, & James, 2007;

Motowidlo, Hooper, Jackson, 2006) while self-reported measures do (Cook, 1993; Hogan, Hogan, & Roberts, 1996). Therefore, when researchers and practitioners use self-report measures, they find that socially desirable responses are not valid for assessing one's personality, especially when they are looking for non-aggressive employees. For instance, one study by Rosse, Stecher, Miller, and Levin (1998) found a significant difference between the neuroticism scores of job applicants and those of job incumbents. As job applicants want to impress their prospective employers, they tend to respond in a socially desirable way while job incumbents who already have a job are less likely to do so.

LeBreton et al. (2007) also investigated faking issues associated with the CRT-A. In one of their three studies, they revealed the purpose of the CRT-A: to identify individuals who are unconsciously ready to justify their aggressive tendencies. Once they became aware of the purpose of the test, the respondents were able to select aggressive responses, so their scores were significantly higher than those of the control group (i.e., those following the normal instructions). The other group, who were told that the CRT-A assesses underlying personality traits, known as aggression, and were asked to select the most logical response, scored higher than the control group. Even though the participants in this group were aware that the test was assessing aggression, they were more likely to choose aggressive alternatives. The results opposed LeBreton et al.'s expectations, but this could be due to the small sample size and the power of suggestion. Furthermore, in another study, LeBreton et al. (2007) compared the mean scores of the CRT-A of undergraduate students, job applicants, and job incumbents. Unlike the self-reported measures, the CRT-A showed no significant differences among the mean scores of the three groups, suggesting that job applicants do not or cannot respond in a socially desirable way.

To investigate more measurement issues (i.e., faking) on the Korean CRT-A, this study will replicate the LeBreton et al. study with Koreans (2007). If the study confirms invariance of scores on the Korean CRT-A, it would indicate that Koreans are able to identify aggressive alternatives from the four options in the CRT-A items. In addition, to remain a resistance-faking measure, the CRT-A should exhibit no significant differences between Korean undergraduates' scores and Korean employees' scores on the test. Thus, this study has formulated the following hypotheses:

Hypothesis 1: *The mean score of experimental group 1 (i.e., instructed to select aggressive responses from the personality measure that appeared to be a reasoning test) will be higher than that of the control group. The mean score of experimental group 2 (i.e., instructed to select the most logical response from the personality measure that appeared to be a reasoning test) will be lower than that of the control group.*

Hypothesis 2: *The CRT-A scores of Korean undergraduates will not significantly differ from those of Korean job incumbents.*

CHAPTER 5

METHODOLOGY

Study 1 (Equivalence of the AQ and the CRT-A)

Participants

Korean Participants

Four hundred and six students enrolled in universities in Korea participated in this study. After excluding participants who had lived in foreign countries for more than three years and participants who endorsed more than five illogical alternatives (James & McIntyre, 2000), 363 participants remained. The mean age of the final sample was 20.02 and 40.5% were male.

US Participants

Five hundred and sixty-four American students who were enrolled in a psychology course were recruited for this study. Just as for Korean participants, American students who primarily resided in foreign countries and students who endorsed more than five illogical alternatives were dropped from further analysis. Remaining were 432 students; their mean age was 19.5, and 55.6% were male.

Procedure

Translation

The most popular translation process is back-translation, which has shown to be successful since the 1960s (Fink, 1963; Werner & Campbell, 1970; Sinaiko, 1963). The author of this study, whose native language is Korean and who is familiar with the CRT-

A, translated the original measure into Korean. In addition, to enhance the reliability of the Korean CRT-A for native Korean speakers, a Korean college professor was asked to review the Korean CRT-A and the AQ. Then, a third person, completely unfamiliar with the English CRT-A measure and blind to the purpose of the study, was asked to back-translate it into English. Finally, a native English-speaking psychology student familiar with the CRT-A was asked to check the equivalency of the meanings in the original version of the CRT-A and the back-translated version. Any discrepancies found between the original and Korean versions of the CRT-A were resolved.

Conditional Reasoning Test for Aggression

Implicit aggression was measured using the new CRT-A, which includes five more items than the original version. This test consists of 30 reasoning items including three bogus items. For each item, premises and reasoning tasks are followed by four possible solutions (alternatives). Different scoring systems can be used for the CRT-A (i.e., dichotomous or trichotomous), and this study adapted a dichotomous scoring system. Aggressive alternatives were scored +1, and pro-social and illogical responses were scored 0. High scores indicated highly aggressive personalities, while low scores indicated pro-social personalities.

Table 2 presents a sample item. In this question, alternatives (a) and (c) are illogical responses. The pro-social alternative from the sample item is (b): “It offers no way to settle a conflict in a friendly manner,” and the aggressive alternative is (d): “People have to wait until they are attacked before they can strike,” which is based on the retribution bias. Implicitly aggressive individuals are more interested in seeking

retaliation than in seeking ways to maintain a relationship. From an aggressive individual's perspective, the "eye for an eye" approach is problematic because of the need to wait to attack others, rather than resolving the issue in a friendly manner. As the retribution bias is embedded in the cognitive processes of unconsciously aggressive individuals', they think their beliefs are reasonable and sound; thus, they justify their belief in retribution.

Table 2. Illustrative Conditional Reasoning Problems

-
1. The old saying, "an eye for eye," which means that if someone hurts you, then you should hurt them back. If you are hit, then you should hit back. If someone burns your house, then you should burn their house

Which of the following is the biggest problem with the "eye for eye" plan?
 - a. It tells people to "turn the other cheek."
 - b. It offers no way to settle a conflict in a friendly manner.
 - c. It can be used only at certain times of the year.
 - d. People have to wait until they are attacked before they can strike.
-

Sources: James, R. L., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., Frost, B. C., Russell, S. M., Mitchell, T. R. & Williams, L. J. (2005). A Conditional Reasoning Measure for Aggression. *Organizational Research Methods*, 8, 69-99

Aggression Questionnaire

To measure self-reported aggression, I adopted the 29-item Buss and Perry Aggression Questionnaire, using a 7-point Likert scale. This measure was also translated into Korean following the back-translation technique.

Data Analysis

This study evaluated measurement invariance of the AQ and the CRT-A by applying factor analysis from structural equation modeling and Differential Item Functioning (DIF) from item response theory. Because the AQ is based on categorical variables and the CRT-A is based on binary variables, different factor analyses were used.

Midpoints and extreme points of the AQ scale

According to the AQ scoring system, items 7 and 18 were reverse scored. Then, to compare mean number of American and Korean respondents selecting both midpoints and extreme points of the AQ scales, an independent sample t-test was applied.

Factor Analysis of the AQ

Previous research has shown that the AQ is a four-factor structure; physical aggression, verbal aggression, anger, and hostility (Buss & Perry, 1992). To confirm the four-factor structure, confirmatory factor analyses (CFA) were carried out for both participant groups separately. First, inter-item polychoric correlation matrices obtained from LISREL (Jöreskog & Sörbom, 1996) were entered for CFA using Mplus (Muthen & Muthen, 2006). Factor loadings of each variable (item) were free to be estimated except for the first variable of each factor, which was at 1; factor intercorrelations were

estimated. The comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) were used as fit indices. To be considered a good model fit, the CFI needed to be greater than 0.90, the TLI needed to be greater than 0.95, and the RMSEA needed to be less than 0.06 (Bollen & Long, 1993). EFA was conducted with the maximum likelihood method and promax rotation for both American and Korean participants. A Kaiser-Gutman eigenvalue greater than 1 (Kaiser, 1960) was used to determine a number of factors to be retained. The factor analysis showed that first factor accounted for around 30% of variance, which meant the IRT model was applicable.

Differential Item Functioning (DIF) of the AQ

For DIF analysis, responses were dichotomized, such as four extreme points (1, 2, 6, and 7) to 1, and three midpoints (3, 4, and 5) to 0 (see Tsutsumi et al., 2009). This was done because this study is primarily interested in response patterns of Americans and Koreans, such as whether any significant difference exists in using extreme points (extremely uncharacteristic of me or extremely characteristic of me) or midpoints. Furthermore, displaying 7-point categorical data is too complex for an item characteristic curve, and the complex graph does not provide much information.

BILOG-MG software (du Toit, 2003) was used to conduct DIF analysis. The American group was assigned as a reference group, and the Korean group was set as a focal group. This study followed the recommendations of Thissen, Steinbert, and Weiner (1993, 1998) regarding the IRT likelihood ratio model's use to detect DIF items. The likelihood ratio model suggests that if the values of -2 times the log-likelihood for the

augmented model are significantly greater than -2 times the log-likelihood for the baseline model, then at least one item displays DIF. Each item was evaluated based on the assumption that “a difference between thresholds” greater than 0.3 means that DIF exists in the item (Tsutsumi et al, 2009).

Factor Analysis of the CRT-A

Principal axis factoring using a tetrachoric correlation matrix was conducted with promax rotation for an American sample. A number of factors were determined based on an eigenvalue greater than 1. One of the CRT-A items, CRT-A 7, had a very low response rate; only seven participants, out of 432, endorsed aggressive responses. This item had almost no variance between items and, thus, was dropped for further factor analyses. To confirm the American CRT-A factor-structure, CFA with Korean students was conducted. The same fit indices used in the previous analysis-CFI, TLI, and RMSEA-were applied to determine the model fit. Just as with the AQ factor analysis, the principal axis factoring results showed that the IRT model is applicable.

Differential Item Functioning (DIF) of the CRT-A

The CRT-A was scored dichotomously and, therefore, recoding was not necessary. The remaining steps of DIF analysis followed the same steps of the AQ DIF analysis.

CHAPTER 6

RESULTS AND DISCUSSION

Results

Mean Number of Respondents Using the Midpoint and Extreme Points (1 or 7)

An independent sample t-test was conducted to compare the mean number of respondents selecting the midpoint (4 = neither uncharacteristic nor characteristic of me) for Koreans and Americans. No significant difference in mean number using midpoints was found for Koreans ($M = 3.06$, $SD = 2.22$) and Americans ($M = 2.96$, $SD = 2.28$); $t(793) = 0.606$, $p > .01$ (Table 3). These results suggest that Koreans did not use the midpoint on the AQ more than the Americans did. On the other hand, a significant difference was found in the mean number of using extreme points (1 = extremely uncharacteristic of me, 7 = extremely characteristic of me) between Koreans ($M = 4.95$, $SD = 4.3$) and that of Americans ($M = 7.57$, $SD = 5.34$); $t(791) = -7.52$, $p < .01$. The effect size was large, at *Cohen's d* = .540. The results showed that Americans used more endpoints on a Likert type scale than Koreans did, which is consistent with previous studies (Chen, Lee, & Stevenson, 1995). Response patterns of Korean and American were significantly different; therefore an American's 7 (strongly agree) and a Korean's 7 might not mean the same thing. Considering Korean culture, to use an end point, Koreans must agree or disagree extremely strongly with the statement. However, Americans use those same endpoints when they strongly agree or disagree with the item more literally.

Table 3. Descriptive Statistics for the Aggression Questionnaire in Study 1

Response	Group	N	M	SD
Number of Using Midpoints (4)	Korean	363	3.06	2.22
	American	432	2.96	2.28
Number of Using Extreme Points (1 or 7)	Korean	363	4.95	4.31
	American	432	7.57	5.34

Factor Analyses of the AQ

The AQ is known to assess four different types of aggression. For example, items 1 to 9 were grouped to assess physical aggression, items 10 to 14 measure verbal aggression, items 15 to 21 assess anger, and items 22 to 29 were clustered to assess hostility. Based on this four-factor structure, a CFA was carried out for Americans and Koreans separately. The fit indices indicated that the four-factor structure model did not fit with Americans or Koreans. With Americans, the CFI was .770, TLI was .749, and RMSEA was .102 (Table 4). With Koreans, the CFI was .710, TLI was .683, and RMSEA was .096. A modified model with correlated measurement error terms did not improve the model fit. With these data, I was not able to replicate the same four-factor structure for either sample. Therefore, to explore a factor structure of the AQ, EFA was conducted separately for Americans and Koreans.

Table 4. Results of Confirmatory Factor Analysis for the Aggression Questionnaire.

Variable	Americans	Koreans
n	432	363
chi-square	2030.13	1607.79
<i>df</i>	371	371
CFI	.77	.71
TLI	.75	.68
RMSEA	.10	.10
90 % CI for RMSEA	0.10-0.11	0.09-0.10

Note. CFI = Comparative fid index; TLI = Tucker-Lewis index; RMES = root mean square error of approximation; CI = confidence interval

For EFA with promax rotation and a Kaiser-Gutman eigenvalue greater than 1 criterion, Americans showed a six-factor structure (Table 5). In this six-factor structure, all 29 items had relatively high factor loadings, ranging from .383 to .844. Items 1, 2, 4, 6, 8, and 9 loaded highest on Factor 1; items 3, 5, and 7 loaded on Factor 2; items 10, 11, 12, 13, 14, and 15 on Factor 3; items 16, 17, 18, 19, 20, and 21 on Factor 4; items 22, 23, 24, 25, 26, and 28 on Factor 5; and items 27 and 29 on Factor 6. Three items from “physical aggression” (Items 3: “If somebody hits me, I hit back”; Item 5: “If I have resort to violence to protect my rights, I will”; and Item 7: “I can think of no good reason for ever hitting a person”) were shown as a separate factor, and two items from “hostility” (Item 27: “I am suspicious of overly friendly strangers”; and Item 29: “When people are especially nice, I wonder what they want”) were not clustered with other hostility items. Thus, compared to the original four-factor structure, two extra factors

were extracted, one from physical aggression and one from hostility. The two new factors, Factor 2 and Factor 5, were labeled as “relational physical aggression” and “distrust of friendliness,” respectively.

Next, another EFA was conducted with the Korean sample. AQ Item 14 showed that its loadings were greater than 1 (which was inferred in Heywood cases [Dillon, Kumar, & Mulani, 1987]; details of the Heywood case are discussed below), and it was dropped from further analysis. Based on the eigenvalue greater than 1 criterion, the same number of factors, six, was extracted. However, compared with American sample, different items were grouped together from physical aggression and relational physical aggression (Table 6). For example, two items from physical aggression (Item 4 and Item 6) were clustered separately, and Item 13 showed split loadings between physical and verbal aggression. For those items, 4 and 6, in the Korean AQ, I used the word “fight” (싸우다 in Korean), which might seem ambiguous to Koreans; it could mean physical or verbal aggression. In English, the phrases “got into fights” and “we came to blows,” literally mean a physical fight. But the word “fight” in Korean could mean physical aggression to some people and verbal aggression to others. Koreans demonstrated good replications for the remaining four (verbal aggression, anger, hostility, antagonism of friendliness) out of six factors.

Table 5. Factor Loadings (EFA) on the Aggression Questionnaire for Americans

AQ Item	F1 (Hostility)	F2 (Physical Aggression)	F3 (Verbal Aggression)	F4 (Verbal and Physical Aggression)	F5 (Distrust of Friendliness)	F6 (Anger)
AQ1	0.821					
AQ2	0.639					
AQ3		0.629				
AQ4	0.642					
AQ5		0.683				
AQ6	0.753					
AQ7		0.678				
AQ8	0.68					
AQ9	0.447					
AQ10				0.423		
AQ11				0.626		
AQ12				0.411		
AQ13				0.889		
AQ14				0.926		
AQ15				0.389		
AQ16						0.466
AQ17						0.538
AQ18						-0.583
AQ19						0.596
AQ20						0.785
AQ21						0.831
AQ22			0.613			
AQ23			0.76			
AQ24			0.779			
AQ25			0.722			
AQ26			0.374			
AQ27					0.731	
AQ28			0.47			
AQ29					0.839	

Table 6. Factor Loadings (EFA) on the Aggression Questionnaire for Koreans.

AQ Item	F1 (Hostility)	F2 (Physical Aggression)	F3 (Verbal Aggression)	F4 (Verbal and Physical Aggression)	F5 (Distrust of Friendliness)	F6 (Anger)
AQ1		0.404				
AQ2		0.673				
AQ3		0.736				
AQ4				0.595		
AQ5		0.65				
AQ6				0.376		
AQ7		-0.399				
AQ8		0.33				
AQ9		0.357				
AQ10			0.798			
AQ11			0.534			
AQ12			0.559			
AQ13			0.35	0.394		
AQ14						
AQ15						0.488
AQ16						0.652
AQ17						0.735
AQ18						-0.493
AQ19						0.278
AQ20						0.433
AQ21						0.543
AQ22	0.401					
AQ23	0.645					
AQ24	0.596					
AQ25	0.643					
AQ26	0.699					
AQ27					0.856	
AQ28	0.737					
AQ29					0.648	

DIF Analysis of the AQ

The IRT likelihood ratio model indicated that the DIF model fit better than the non-DIF model, meaning that -2 times the log of the likelihood (G^2) of the augmented model was significantly greater than -2 times the log of the likelihood of the baseline model. G^2 of the augmented model was 27860.8209, and G^2 of the baseline model was 27455.7215. The results indicated that at least one item functioned differently across cultures. Item level analysis showed that 20 out of 29 items showed DIF for Korean samples (Table 7). The results indicate that Americans' difficulty levels of selecting extreme points for the 20 items of the AQ are different from Koreans' levels. An Item Characteristics Curve (ICC) suggested that ICCs for American and Korean AQ 9 and 17 (Figures 1 and 2) are quite similar, as no DIF exists in these items. Americans and Koreans showed comparable trait levels for selecting extreme points on the AQ 9 and 17. However ICCs of AQ 19 and 24 (Figures 3 and 4; DIF exists) are not comparable. This suggests huge differences in latent aggression levels in the two groups.

For example, Koreans felt it more easy to use extreme points on Item 19 ("Some of my friends think I'm a hothead") than did Americans. Typically, Koreans are less likely to express their feelings about others in a direct way, and "You are a hothead" is a very uncommon expression in Korea. Thus, Koreans tend not to think that others would think, "I am a hothead"; thus, they felt comfortable using extreme points, especially 1 (extremely uncharacteristic of me) or 2, on this item. On the other hand, Koreans found it more difficult to use extreme points on Item 24 ("Other people always seem to get the breaks") than did Americans. DIF may have occurred on this item because its wording seems to be vague and unclear to Koreans. In Korean, if an item does not specify a

particular occasion, readers might be confused and would not have a clear idea about the item's meaning. This item seemed to confuse Koreans, and thus, they found it particularly difficult to use extreme points in response. These DIF results are consistent with the first results in this study that the response patterns of Koreans and Americans are significantly different.

Table 7. IRT adjusted threshold parameters of the Aggression Questionnaire items between Korean and American

AQ Item	b_r (American)	b_f (Korean)
AQ1	-1.994	-3.140*
AQ2	-0.453	-1.029*
AQ3	0.275	-0.044*
AQ4	-2.231	-3.266*
AQ5	0.329	-0.210*
AQ6	-0.453	-1.959*
AQ7	-0.120	-0.279
AQ8	-0.603	-1.589*
AQ9	-0.604	-0.616
AQ10	0.735	0.850
AQ11	0.833	0.994
AQ12	0.196	0.486
AQ13	-1.279	0.120*
AQ14	-0.958	-0.598*
AQ15	-0.081	0.264*
AQ16	0.329	0.628
AQ17	-1.062	-0.987
AQ18	0.680	0.802
AQ19	-0.279	-1.412*
AQ20	-1.154	-1.340
AQ21	-0.828	-1.050
AQ22	0.355	-0.060*
AQ23	-1.263	0.055*
AQ24	-1.018	0.581*
AQ25	0.025	0.120
AQ26	-0.659	0.136*
AQ27	-0.081	0.264
AQ28	-0.549	0.055*
AQ29	-0.094	0.216*

Note. b_r : adjusted threshold parameters of reference group (American); b_f : adjusted threshold parameters of focal group (Korean). * indicates existence of DIF.

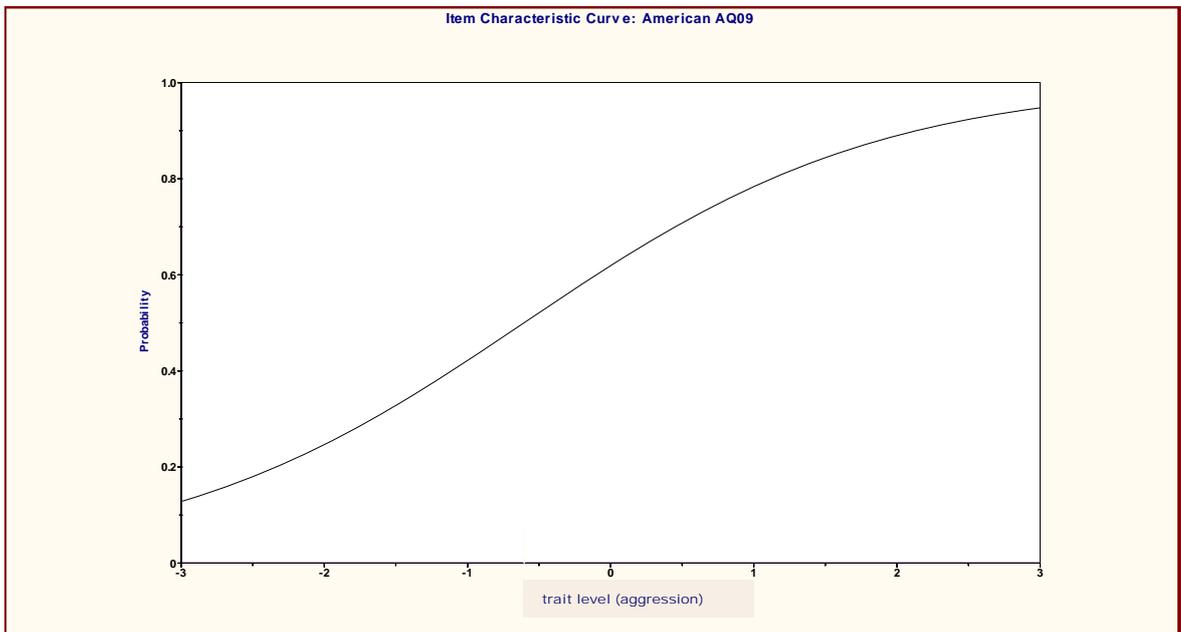
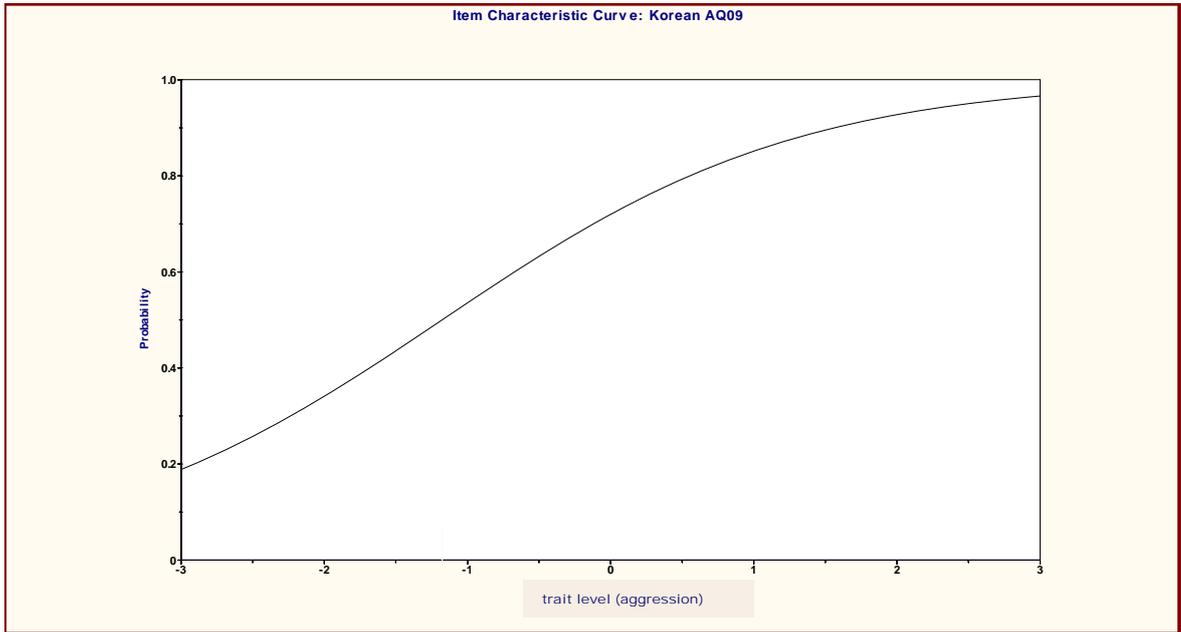


Figure 1. Item Characteristic Curves for Aggression Questionnaire Item 9: Korean versus American

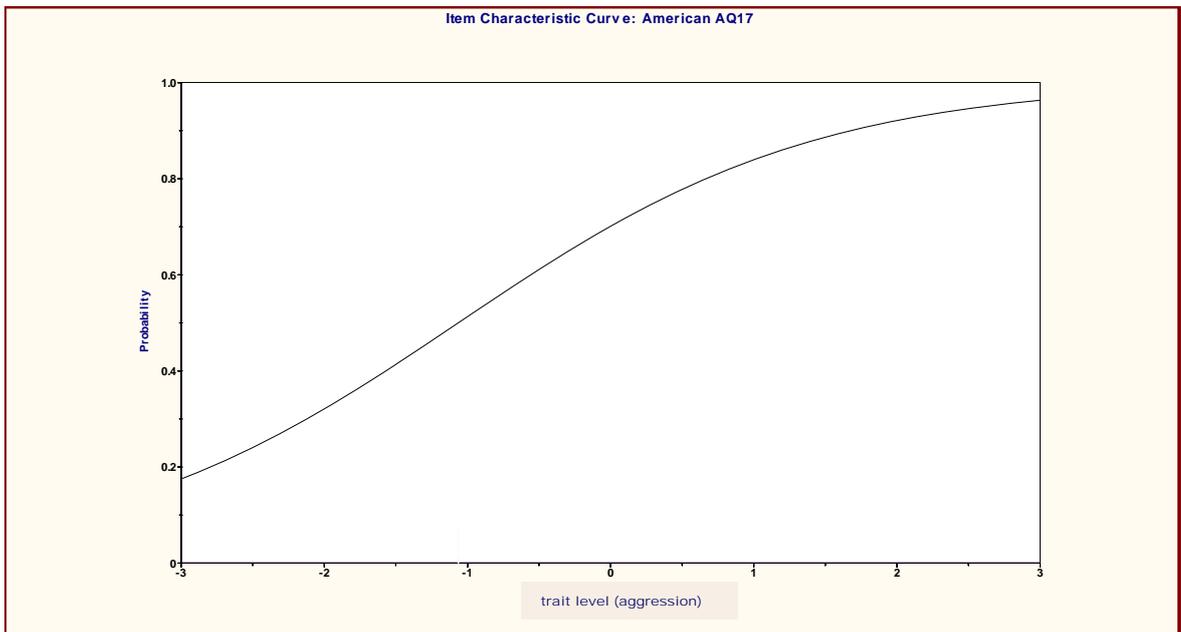
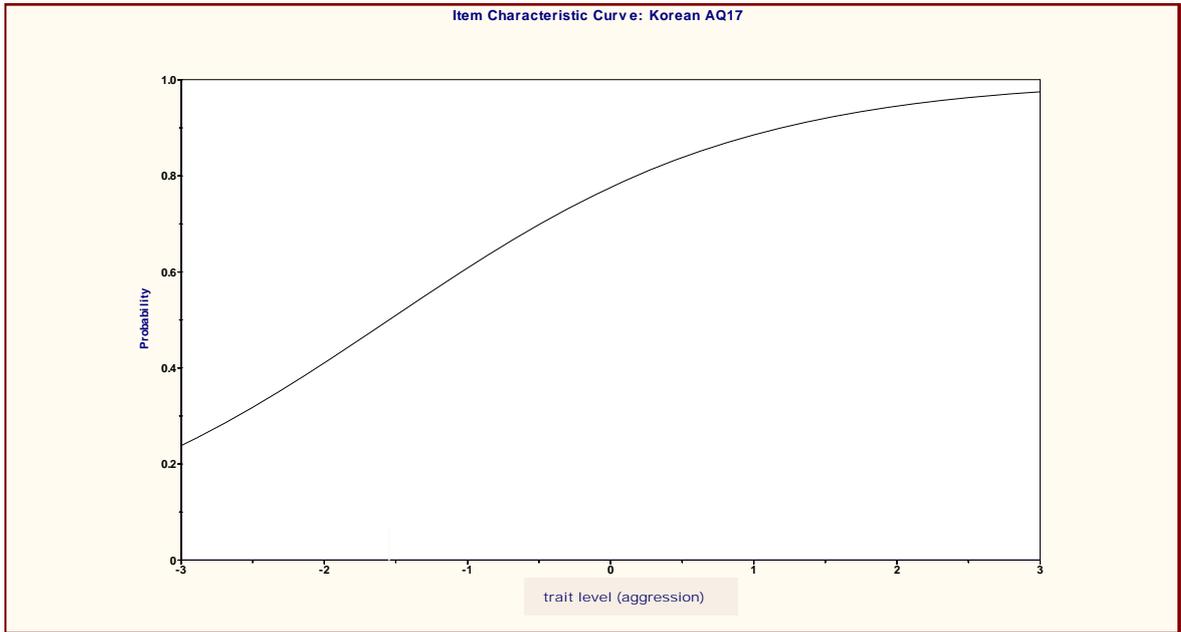


Figure 2. Item Characteristic Curves for Aggression Questionnaire Item 17: Korean versus American

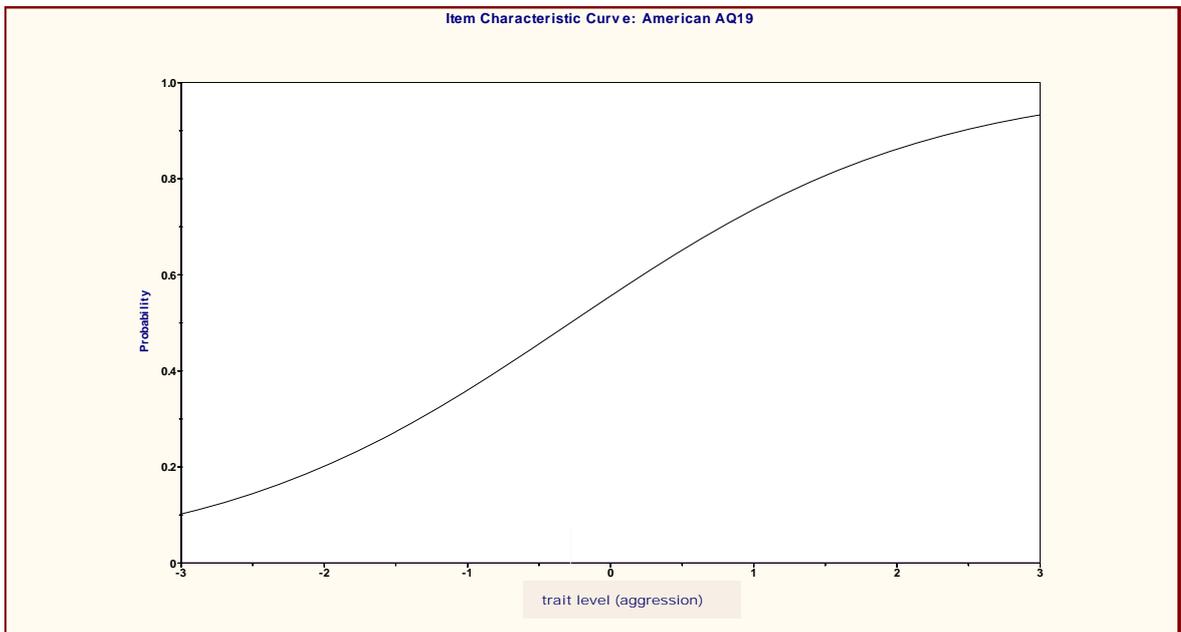
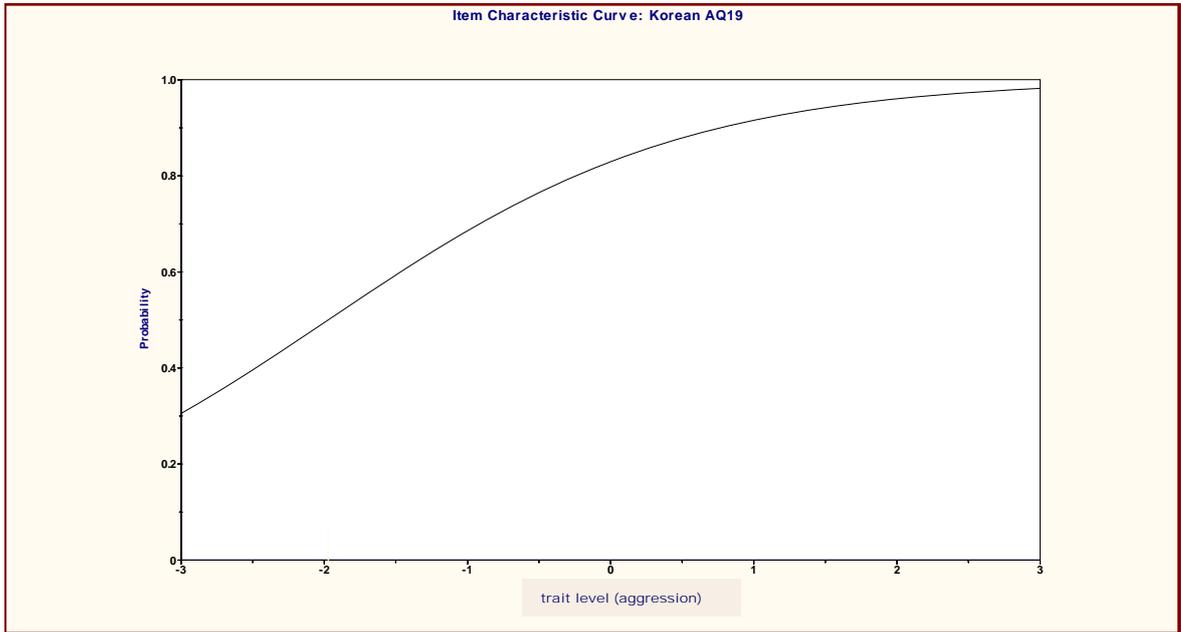


Figure 3. Item Characteristic Curves for Aggression Questionnaire Item 19: Korean versus American

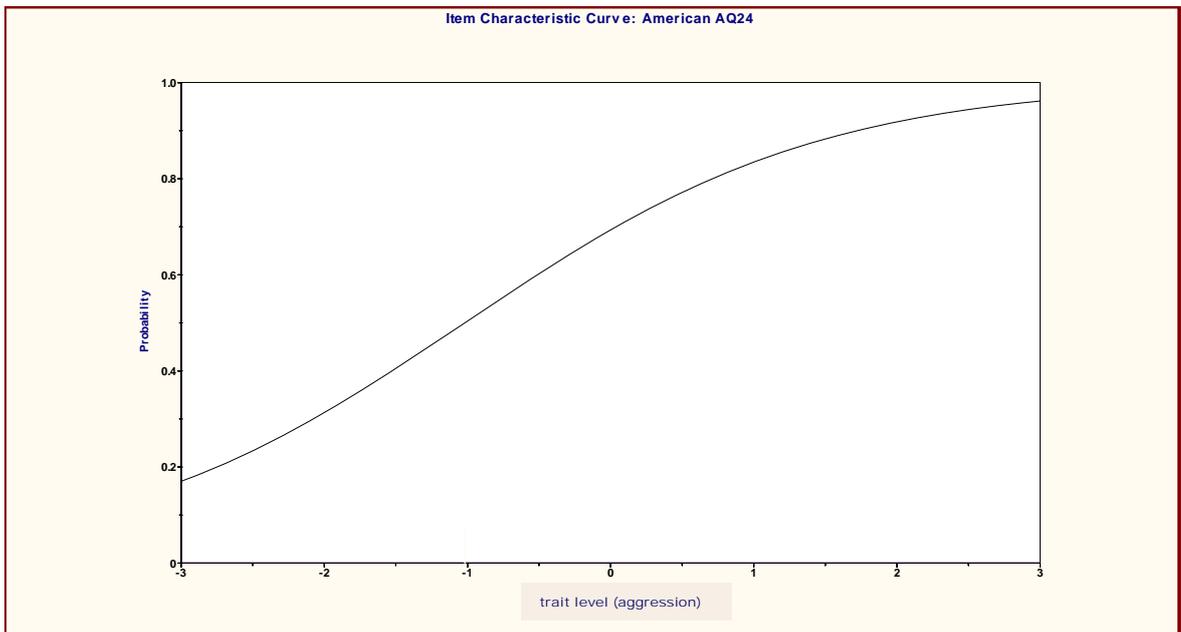
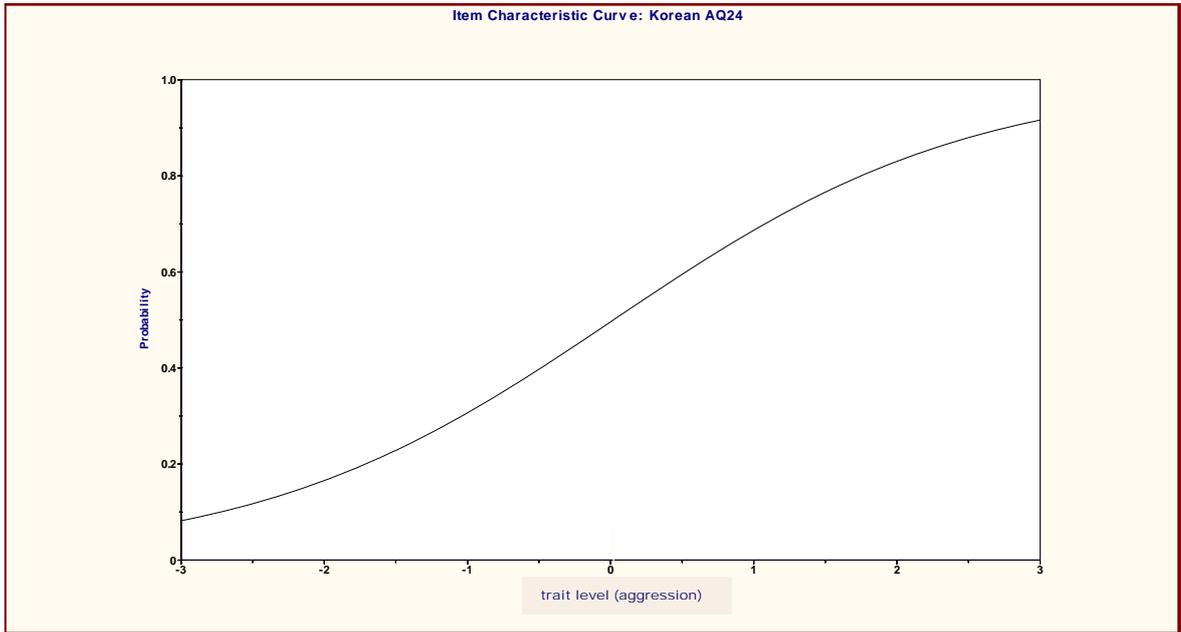


Figure 4. Item Characteristic Curves for Aggression Questionnaire Item 24: Korean versus American

Factor Analyses of the CRT-A

The principal factor axis using promax rotation with American CRT-A data is presented in Table 8. An eigenvalue greater than 1 criterion showed a four-factor structure, which was a little different from the CRT-A three-factor structure (James & LeBreton, 2011). The difference could be caused by adding five new items and perhaps sampling error due to the relatively small number of the students (N = 432) compared to the sample size for the three-factor structure (N = 4772). The four factors accounted for 76% of the total variance. Five CRT-A items loaded highest on Factor 1, eight items loaded on Factor 2, five items on Factor 3 and seven items on Factor 4. Compared to the three-factor structure, external controls (Factor 4) and internal controls (Factor 1) were moderately replicated. Based on the six JMs of the CRT-A, Factor 2 and Factor 3 were labeled as “hostility of powerful others” and “potency,” respectively.

To confirm the four-factor structure of the CRT-A, CFA was supposed to be run with the Korean samples. However, because of a singular matrix, which caused a non-positive definite (Heywood cases¹) CFA could not be carried out.

¹ Heywood cases occur when there are non-positive definite matrices, high multicollinearities, linear dependencies, minor data entry problems (i.e., typographical errors), large amounts of missing data, misspecified (underidentified) models, small samples, and/or outliers, which can lead to negative variance estimates (Brown, 2006; Dillon, Kumar, & Mulani, 1987). Based on CFA with Mplus and EQS, Heywood cases with the CRT-A data seemed to be due to a non-positive definite correlation matrix, an

Thus, to understand the factor structure of the Korean CRT-A, another principal factor axis analysis was conducted. With the Korean sample one more factor was extracted, which represents a five-factor structure (Table 9). Three Korean CRT-A items loaded highest on Factor 1, three items on Factor 2, seven items on Factor 3, four items on Factor 4, and nine items on Factor 5. The different number of factors and different pattern of factor structure indicates that the CRT-A may not assess implicit aggressiveness among Koreans in the same way that it assesses Americans implicit aggressiveness. There may be construct bias, and this failure to replicate the factor structure of the CRT-A could be due to different cultural issues and/or due to the tetrachoric correlation matrix with binary data. Embretson and Reise (2000) mentioned that, "Tetrachoric correlations are preferred over phi correlations because they correct for item difficulty effects... Adjusting whole matrix of item correlations to tetrachorics sometimes results in a singular correlation matrix, which is not appropriate for factor analysis" (p.37). As mentioned above, a singular matrix was the case in this study, wherein a tetrachoric correlation matrix was entered, thus, factor analysis of the CRT-A may not provide meaningful information. Therefore, just as with the AQ invariance process, DIF analysis from the IRT model was used.

underidentified model, linear dependencies, and a relatively small sample size. Remedies for Heywood cases was not applicable with these data and "quick remedies (i.e. setting error variance to zero)" could cause other problems (Brown, 2006).

Table 8. Factor Loadings (EFA) on the Conditional Reasoning Test-Aggression for Americans.

CRT-A Item	F1	F2	F3	F4
CRT-A 3				-0.198
CRT-A 4	-0.137			
CRT-A 5	-0.764			
CRT-A 8	0.982			
CRT-A 9			0.290	
CRT-A 10		0.592		
CRT-A 11			0.174	
CRT-A 12		0.257		
CRT-A 13			0.517	
CRT-A 14				0.342
CRT-A 15				0.273
CRT-A 16		0.210		
CRT-A 17				
CRT-A 18		0.700		
CRT-A 19	0.217			
CRT-A 20		0.282		
CRT-A 21			0.294	
CRT-A 22				-0.223
CRT-A 23				0.184
CRT-A 24		0.448		
CRT-A 25				0.518
CRT-A 26	0.377			
CRT-A 27			-0.312	
CRT-A 28		0.441		
CRT-A 29				0.368
CRT-A 30		0.466		

Table 9. Factor Loadings (EFA) on the Conditional Reasoning Test-Aggression for Koreans.

CRT-A Item	F1	F2	F3	F4	F5
CRT-A 3					0.286
CRT-A 4					-0.294
CRT-A 5	-0.699				
CRT-A 8					0.616
CRT-A 9					0.270
CRT-A 10			0.457		
CRT-A 11				0.257	
CRT-A 12		0.197			
CRT-A 13					-0.173
CRT-A 14					0.282
CRT-A 15				0.508	
CRT-A 16	0.958				
CRT-A 17				0.361	
CRT-A 18			0.323		
CRT-A 19					-0.362
CRT-A 20	-0.823				
CRT-A 21		0.828			
CRT-A 22		-0.957			
CRT-A 23			0.209		
CRT-A 24			0.409		
CRT-A 25			0.369		
CRT-A 26					0.228
CRT-A 27				0.329	
CRT-A 28			0.458		
CRT-A 29					0.176
CRT-A 30			0.596		

DIF Analysis of the CRT-A

Again, the IRT likelihood ratio model indicated that the DIF model fit was better than the non-DIF model. G^2 of the augmented model was 6930.423 and G^2 of the baseline model was 6236.5613, which indicates at least one item showed DIF in the CRT-A. The Koreans' difficulty in endorsing an aggressive alternative is different from that of the American participants. Item level analysis suggested that threshold differences between Koreans and Americans were greater than .3 for 26 items, meaning that DIF existed in almost all items on the CRT-A (Table 10). Only one item, CRT-A 24, did not show DIF; thus, its ICCs were similar across groups (Figure 5), while the CRT-A 18 ICC of reference (Americans) and focal (Koreans) groups was quite different (Figures 6, 7, and 8 for CRT-A 16, 18, and 25, respectively). From the Koreans' trait level, selecting an aggressive alternative on CRT-A Item 18 seemed to be easier than it was for the Americans' trait level. Strong DIF on Item 18 could be due to the wording effect (Wu, 2008). The back-translated pro-social alternative was "Hardworking employees receive bonuses and some time off." In the pro-social alternative for the original CRT-A, "bonuses" was the subject of the sentence, but in the Korean CRT-A, "hard-working employees" was the subject of the sentence. Thus, to Koreans it seemed too obvious that employees who work hard receive bonuses, and they were less likely to think of an alternative reason for companies to use bonuses. Consequently, the pro-social option was less attractive to Koreans as a logical alternative; they were more likely to choose the aggressive alternative because the other two options did not sound sensible. The wording seemed to lead Koreans to choose the aggressive alternative more easily, regardless of their aggressiveness.

DIF also occurred in Item 16, which raised concern before data collection. This Item referenced cultural familiarity of U.S. and Japanese carmakers. Accordingly, in the Korean CRT-A, U.S. was changed to Korea and Japan remained the same; however, Koreans tend to have animosity towards the Japanese from a long national history. Thus, apart from each Korean's aggressive tendencies, different cultural attitudes of Koreans and Americans towards Japan seemed to cause DIF on this item.

DIF existed in CRT-A 25, which references World War II, and the aggressive alternative is "Only weak countries follow agreements." Koreans tend to believe that Korea is a weak country while Americans tend to have pride in the US and believe the US is a strong country. Thus, Koreans and Americans are likely to have different perspectives regarding what constitutes a weak country, which seemed to affect DIF in Item 25.

Table 10. IRT adjusted threshold parameters of the Conditional Reasoning Test-Aggression items between Korean and American

CRT-A Item	b_r (American)	b_f (Korean)
CRT-A 3	2.318	7.172*
CRT-A 4	3.872	4.712*
CRT-A 5	7.999	9.409*
CRT-A 7	12.418	9.546*
CRT-A 8	7.783	7.099*
CRT-A 9	1.209	0.802*
CRT-A 10	6.629	2.195*
CRT-A 11	4.039	3.523*
CRT-A 12	1.062	-2.266*
CRT-A 13	4.765	6.070*
CRT-A 14	1.357	0.870*
CRT-A 15	5.121	7.099*
CRT-A 16	5.507	10.697*
CRT-A 17	3.913	4.579*
CRT-A 18	7.680	-0.465*
CRT-A 19	4.124	4.940*
CRT-A 20	3.324	7.988*
CRT-A 21	3.592	5.956*
CRT-A 22	7.580	10.326*
CRT-A 23	2.548	3.448*
CRT-A 24	3.750	3.716
CRT-A 25	9.690	3.077*
CRT-A 26	-0.625	1.175*
CRT-A 27	-2.854	-1.815*
CRT-A 28	6.407	4.155*
CRT-A 29	2.515	0.076*
CRT-A 30	-0.397	1.243*

Note. b_r : adjusted threshold parameters of reference group (American); b_f : adjusted threshold parameters of focal group (Korean). * indicates existence of DIF.

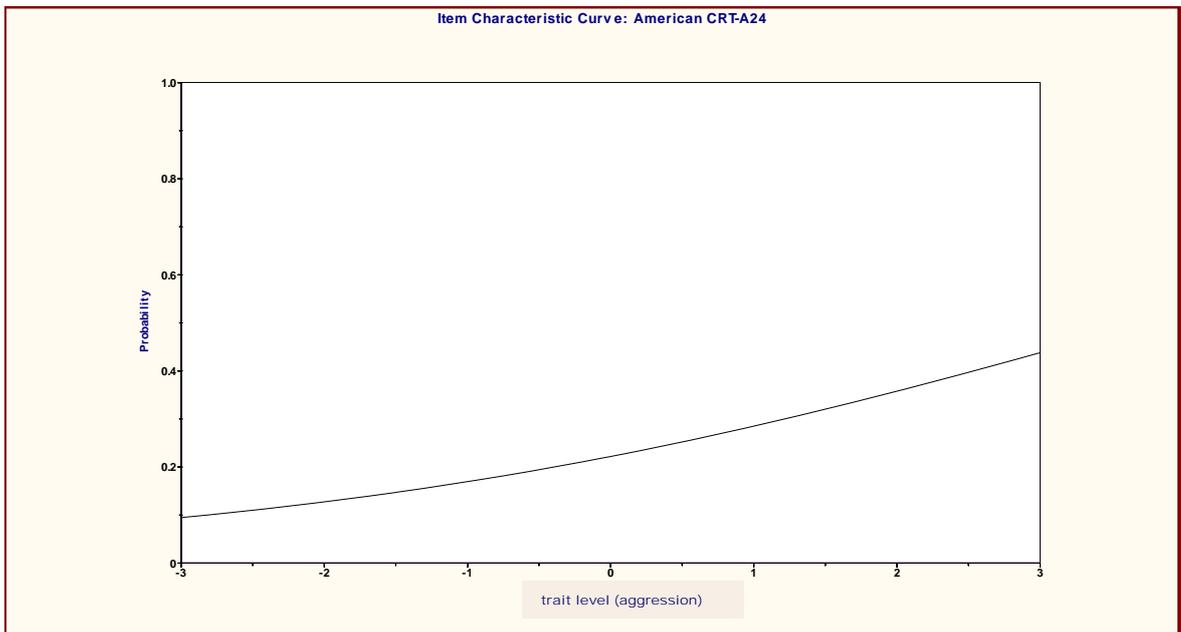
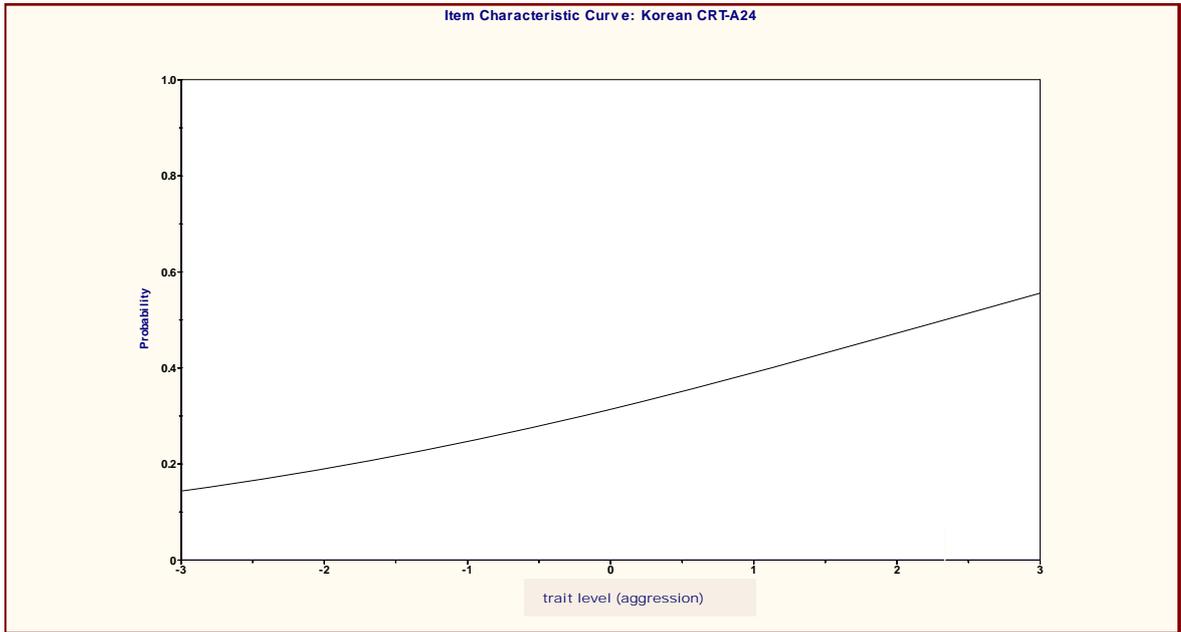


Figure 5. Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 24: Korean versus American

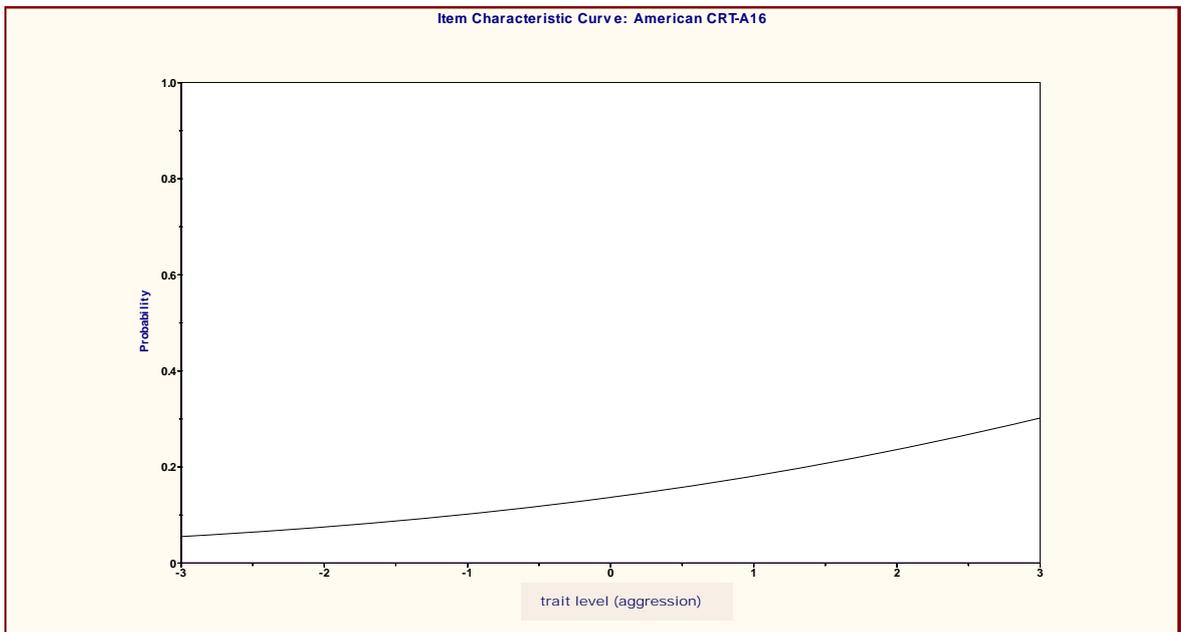
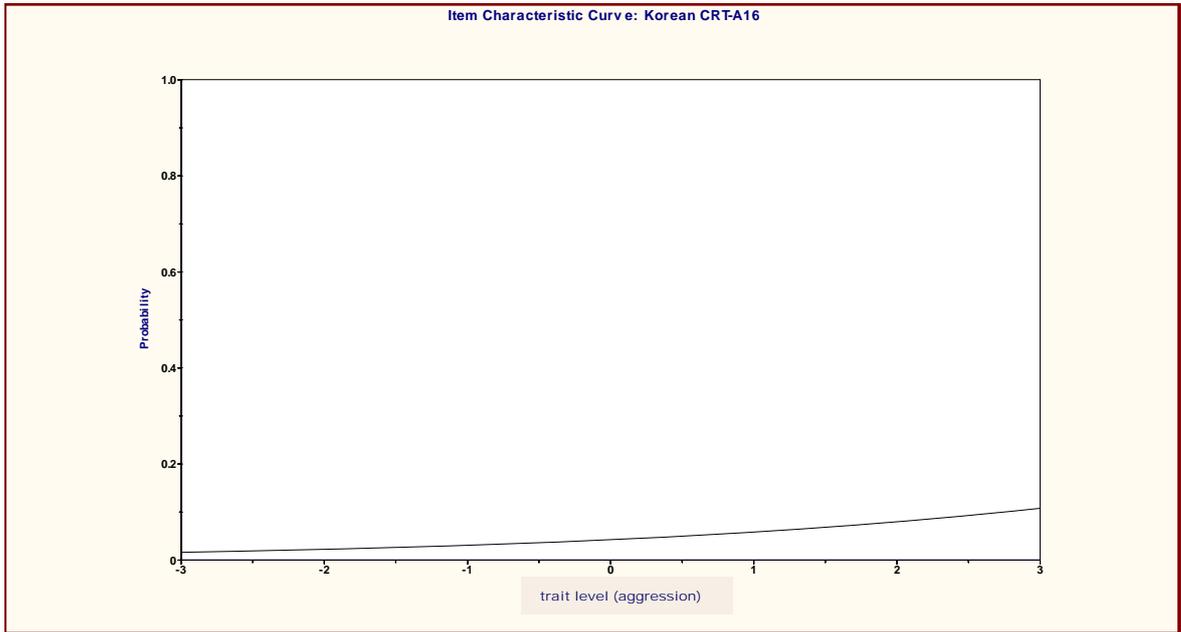


Figure 6. Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 16: Korean versus American

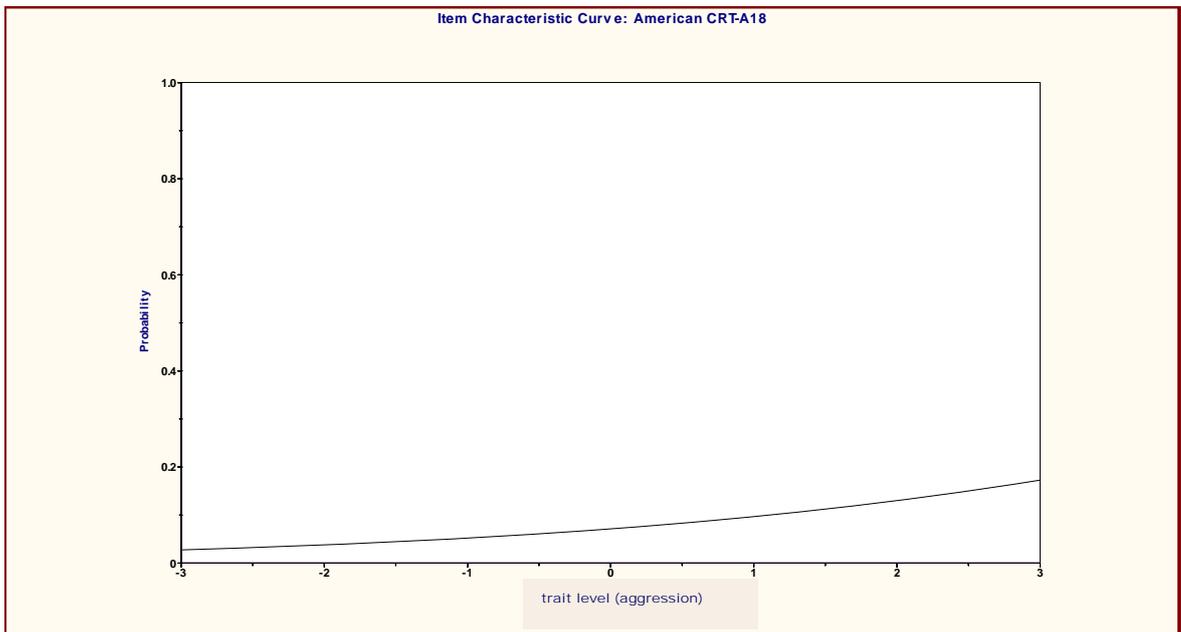
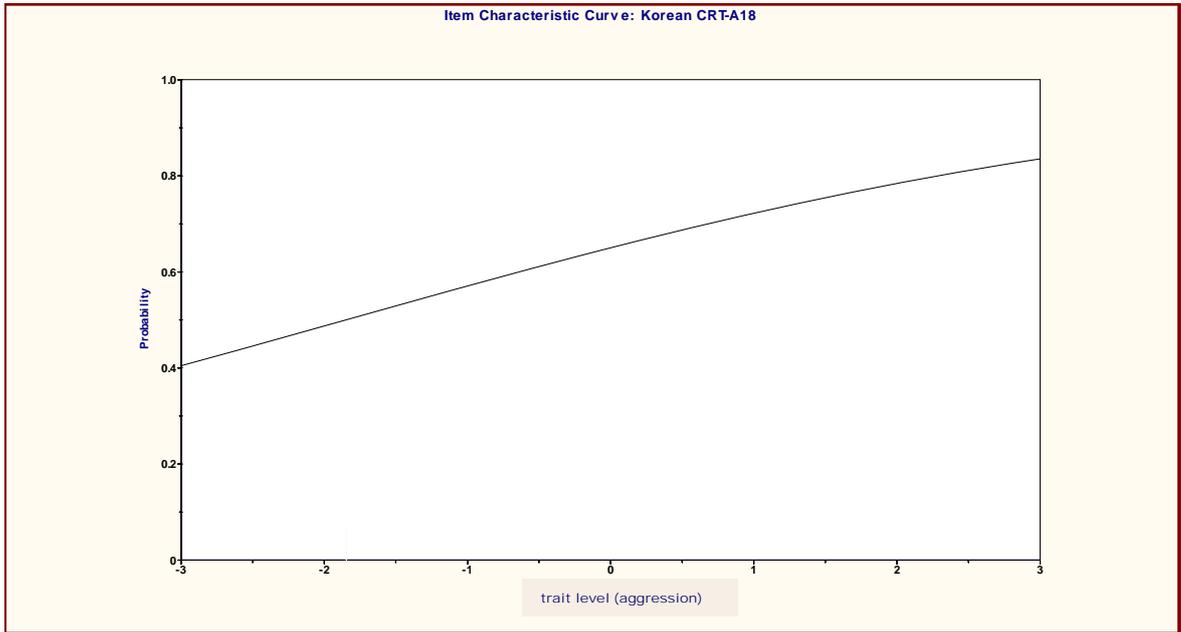


Figure 7. Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 18: Korean versus American

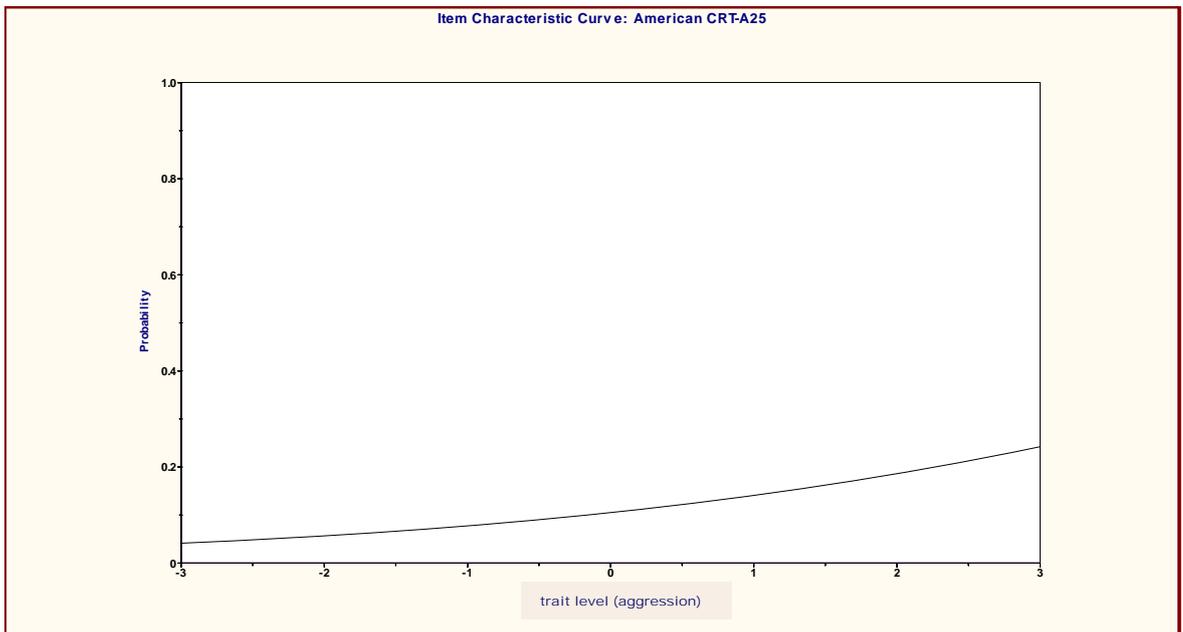
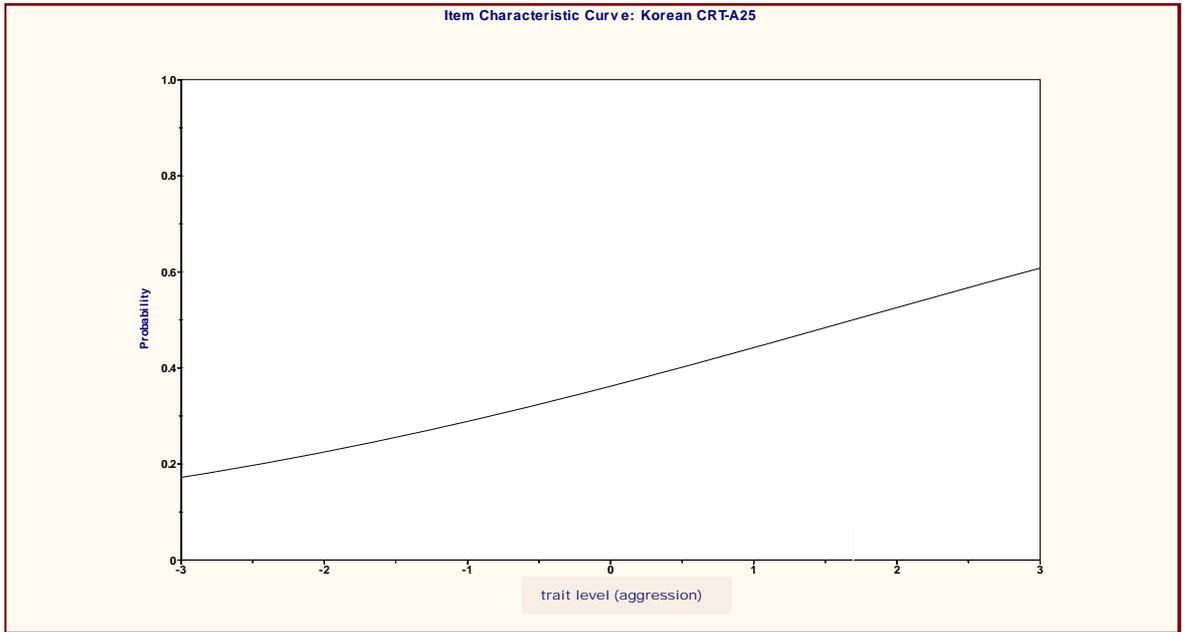


Figure 8. Item Characteristic Curves for Conditional Reasoning Test-Aggression Item 25: Korean versus American

Discussion

Study 1 explored measurement invariance between the two different types of assessments of aggression, the AQ and CRT-A. Construct and item bias seem to exist in both measures when they were applied to samples from different cultures. First, previous studies showed that Asians were not more likely to use midpoints but were less likely to use extreme points than were Westerners. This study supports previous research that Koreans use fewer extreme points and significant differences in their response patterns. Asian children who are likely to be influenced by Confucian philosophy, tend to rate themselves lower on negative personal characteristics (Stevenson et al., 1990). Moreover Asians learn that they should not strongly express their opinions and should be modest (Markus & Kitayama, 1991). This cultural characteristic seems to make Koreans less likely to choose extreme points.

The results support the reference group effect (Heine, Buchtel, & Norenzaya, 2008; Hein et al., 2002), which indicates that when participants respond to a Likert type scale, they tend to compare themselves to the people around them. Therefore, if people from different cultures do not use the same reference group, the seemingly same point on a scale such as 1 point of Americans may not mean the same as a 1 point for Koreans. This suggests that Americans tendency to use more extreme points does not necessarily mean that they are more aggressive. Americans and Koreans just have different response patterns.

A self-reported AQ, which was developed with Americans. did not seem to be equivalent with the Korean AQ. The American AQ was supposed to assess physical aggression, verbal aggression, anger, and hostility. However, with these data, American

college students and Korean college students did not show the same factor structure. American college students showed that there were six subscales under the AQ. Korean college students also showed six-factor structure of the AQ, but the items did not cluster together in the same way as they did for Americans, especially items assessing physical aggression.

Furthermore, DIF analysis showed that Americans and Koreans trait level estimates (aggression) are not comparable. More than half of the items (20 items) of the AQ assessed different levels of aggression between Americans and Koreans. The DIF results are consistent with previous classical testing theory that indicates significant differences in response patterns.

Another type of aggression measure, the CRT-A, which assesses implicit or unconscious level of aggression, did not demonstrate measurement equivalence across cultures. For this study, five new items were developed and tested with Koreans. The four factors were extracted from a new version of the CRT-A with a five-factor structure. Results show that Koreans may have different underlying biases to justify their unconscious motives to be aggressive. However, an inability to run CFA due to a singular matrix and a relatively small sample size makes the statement inconclusive. Similar to the AQ DIF analysis, DIF exists in the CRT-A items. On some of CRT-A items, Koreans found it easier to endorse aggressive alternatives (e.g., Items 3, 15, or 16) while they found it more difficult to choose aggressive responses on other items (e.g., Items 10, 18, or 29).

CRT-A DIF analysis suggests that Koreans' and Americans' trait levels are significantly different; for certain items, Koreans felt easier endorsing aggressive

alternatives while for other items, Americans felt easier selecting aggressive alternatives. It seems that DIF exists in items that have a sizable frequency difference in endorsing aggressive responses. For instance, on CRT-A Item 18, which showed the greatest DIF, the response rate of selecting aggressive alternative was .07% for Americans compared to 64% for Koreans. Furthermore, on CRT-A Item 25, in which DIF occurred, .04% of Americans endorsed the aggressive alternative while 36% of Korean did the same.

CHAPTER 7

METHODOLOGY

Study 2 (Test of Faking)

One of the distinctive features of the CRT-A is its resistance of faking. Unless participants are told that the CRT-A is assessing aggression trait, they do not know that they are being assessed on their aggressiveness (LeBreton et al. 2007). To understand whether Koreans can identify aggressive alternatives or not, Study 2 replicates LeBreton et al.'s (2007) study with Koreans and suggests the following hypotheses.

Hypothesis 1: (A) *The mean score of experimental group 1 (i.e., instructed to select aggressive responses from the personality measure that appeared to be a reasoning test) will be higher than that of the control group.* (B) *The mean scores of experimental group 2 (i.e., instructed to select the most logical response from the personality measure that appeared to be a reasoning test) will be lower than that of the control group.*

Hypothesis 2: *The CRT-A scores of Korean undergraduates will not significantly differ from those of Korean job incumbents.*

Participants

Korean Student Participants

One hundred and twenty-four participants were assigned as a control group, and they were randomly selected from Study 1 (mean age was 20.01; 42.7% were male). Experimental group 1 consisted of 109 students who attended universities in Korea; their mean age was 20.4, and 37% were males. Experimental group 2 consisted of 105 Korean

college students; their mean age was 20.3, and 39% were males. The low proportion of male participants occurred because one of the schools was a women's university.

Participants were randomly selected into the control group, experimental group 1, or experimental group 2.

Korean Job Incumbents

Ninety-seven participants, who were currently employed, participated in this study. The participants' mean age was 31.5, and 60% male.

Procedure

The procedure was the same as that in the LeBreton et al. (2007) study. The control group took the CRT-A following the normal instructions. The participants in experimental group 1 completed the CRT-A with the following instructions: "The CRT-A appears as a reasoning test, but the test actually assesses individuals' underlying personality. Please select an aggressive alternative from each of the CRT-A items." Experimental group 2 took the CRT-A with the following instructions: "The CRT-A appears as a reasoning test, but the test actually assesses individuals' underlying personality. Please select the most logically appealing alternative." For Hypothesis 2, Korean job incumbents completed the CRT-A following normal instructions.

Data Analysis

ANOVA was used to compare the mean scores of the control group to those of the experimental groups, and an independent sample t-test was conducted to compare the mean scores of the college students to those of the job incumbents.

CHAPTER 8

RESULTS AND DISCUSSION

Results

The one-way analysis of variance (ANOVA) and planned comparison was used to detect significant mean score differences between the control group and the experimental groups. There was a significant main effect on the CRT-A scores between the control group and the experimental groups, $F(2, 337) = 318.25, p < .001$ (Table 11). Hypothesis 1 (A) was supported; there was a statistically significant mean score difference on the CRT-A between the control group ($M = 8.88, SD = 2.71$) and the first experimental group ($M = 18.06, SD = 4.35$), $F(1, 232) = 383.26, p < .001$ ($\eta^2 = 0.6239$). When the participants were told the true purpose of the test, they were able to identify aggressive alternatives. Hypothesis 1 (B) was not supported. The mean scores between the control group ($M = 8.88, SD = 2.71$) and the second experimental group ($M = 8.06, SD = 2.44$) were not significantly different, $F(1, 228) = 5.58, p = .019$. Once participants were told the test assessed one's personality, they were not less likely to choose aggressive alternatives.

Hypothesis 2 was supported; there was no significance difference in the CRT-A scores of students ($M = 8.88, SD = 2.71$) and that of employees ($M = 7.97, SD = 2.74$); $t(205) = 2.46, p > .01$ (Table 11). Korean students did not score significantly higher than employees did. The results are consistent with LeBreton et al.'s (2007) findings that there is no significant CRT-A score difference between students and employees for both Koreans and Americans.

Table 11. Descriptive Statistics for the Conditional Reasoning Test-Aggression in Study 2

Group	n	M	SD
Control Group	124	8.88	2.71
Experimental Group 1	109	18.06	4.35
Experimental Group 2	105	8.07	2.44
Job Incumbents	97	7.97	2.74

Discussion

Study 2 revisited LeBreton et al.'s (2007) study to investigate the issue of faking in responses to the Korean CRT-A. The results demonstrated that when the purpose of the CRT-A was fully revealed, participants were able to select the aggressive alternatives, just as shown by English speakers in LeBreton et al.'s study. Koreans in the first experiment group scored significantly higher than Koreans in the control group. The alternatives that appeared aggressive to English speakers also seemed to be aggressive responses to non-English speakers. The second experimental condition was slightly modified from that of the LeBreton et al. In this study, Koreans were told that the CRT-A was a personality survey without telling them what specific personality/trait was being sought, and they were instructed to find the most logical response. Even though they were told that their personality was being assessed, they were not less likely to choose aggressive alternatives. Participants' scores in the second experimental group were not different from the participants' scores in the control group. The results suggest that, even

after they were told they were completing a personality assessment, they did not seem to know the CRT-A was a personality survey that assessed aggressive characteristics. If they had known, their score would have been significantly lower than that of the control group. The idea behind each item in the CRT-A is not as transparent as in the AQ, wherein the purpose of each item is clear to the respondents.

On the Korean CRT-A, there was no significant differences in scores of students and job incumbents. This is consistent with LeBreton et al.'s (2007) study and further supports the CRT-A's resistance to faked responses, unlike self-reported questionnaires which show significant scores differences between students and job incumbents. For instance, Rosse et al.'s (1998) study compared the neuroticism scores of job applicants and job incumbents. Job incumbents' scores were significantly higher than those of job applicants were. Based on the results, job incumbents felt less pressured to fake their responses than did job applicants because job incumbents were already employed, whereas job applicants were actively seeking employment. Thus, the transparency of many personality assessments influences participant responses, especially when gauging an undesirable personality construct (i.e., neuroticism), while items in the CRT-A are less transparent.

CHAPTER 9

GENERAL DISCUSSION

Most psychological assessments validate and test reliability among middle class white samples (Knight, 2000; McLoyd, 1999). Validating the constructs of those assessments with broader populations should provide more information about their psychometric properties. Thus, in this research, the first study investigated measurement invariance of an explicit aggression questionnaire (AQ) and an implicit aggression assessment (CRT-A) across Americans and Koreans using CFA, EFA, and DIF. First, on the AQ, Americans used extreme points considerably more than Koreans. The results show that response patterns of Americans and Koreans are significantly different.

Next, a CFA was conducted based on Koreans and Americans; however, the four-factor structure on the AQ was not replicable. The EFA with Americans suggested six factors, and Koreans showed a similar factor structure for only four subscales: verbal aggression, anger, hostility, antagonism of suspiciousness. A Hong Kong China population also did not support the four-factor structure, but a shorter version of the AQ (12 items) showed a good model fit (Maxwell, 2007). For future study, it would be interesting to run CFA with Koreans on the short form of the AQ. DIF analysis showed that English AQ and Korean AQ are not invariant. DIF existed in two-thirds of the AQ items, suggesting that the difficulty of some items varied between Koreans and Americans. The two groups' latent trait levels were significantly different.

I failed to replicate the factor structure of the English CRT-A with Korean data due to a singular matrix, small sample size, and construct bias. DIF occurred in almost all CRT-A items. Each of the CRT-A items starts with a short premise and those premises seemed much more familiar among American cultures. Therefore, it may not assess Koreans' implicit aggressiveness accurately. The differences on the CRT-A could be cultural differences, translation errors, or different latent variable relationships. The two assessments of aggression need further studies to assess aggressiveness among the Korean population in the same way that it is assessed among Americans. Unless the assessments are developed by researchers from different cultures at the initial stage of development, it may be impossible empirically to meet all the equivalence conditions (van de Vijber & Leung, 2001).

Although the factor analyses and the DIF analysis demonstrated that the CRT-A does not seem to be equivalent for the Korean population, Study 2 results suggested that Koreans could identify aggressive responses when they were told the CRT-A was a personality assessment to identify aggressive individuals. Aggressive alternatives in the CRT-A worked in the same way for both Koreans and Americans.

Even after giving hints to Koreans that the CRT-A was assessing personality traits, they were not able to distort their responses in a socially desirable way. There was no significant difference in the CRT-A scores of the control group and the experimental group 2. When the purpose of the assessment is exposed, participants know what they are being assessed on, and they tend to respond in a socially desirable way especially in a selection process (Rosse et al., 1998). However, in Study 2, even when they were told that they were being assessed on their personality traits their scores were not lower than

the respondents who were not so instructed. Respondents in experimental group 2 may be less motivated to impress others even though they knew that their personality traits were assessed because they were participating in a research study, not a job screening. Participants may have wanted to understand their own personality and answered the questions honestly.

Consistent with LeBreton et al.'s (2007) findings, Korean colleges students and employees did not respond differently on the CRT-A. Unlike the self-reported measure of neuroticism, employees and undergraduate students responded in similar ways. Both groups did not seem to know the purpose of the test; therefore, they did not feel a need to distort their responses.

Contributions and Implications

This study explored measurement (aggression) invariance across cultures using different models: CFA (classical testing theory; CTT) and item response theory (IRT). CFA approaches of measurement invariance are different from IRT approaches because CFA investigates the construct from a scale level while IRT explores it from an item level. Although each approach has its own advantages, Kim, Kim, and Kamphous (2010) argue that only a few studies used both CFA and IRT to study measurement invariance. This study attempted to investigate measurement invariance using both approaches. Although EFA suggested that some of the factors on the AQ and CRT-A were comparable, CFA and DIF approaches suggest that the AQ and the CRT-A are not invariant across cultures; this study opens a door for the next step in understanding latent level differences among American and Korean aggression.

Furthermore, this study adopted both implicit and explicit measures for aggression. This indirect way of assessing aggression is relatively new, and the CRT-A, which is not likely to be correlated with self-reported aggression, provides a new approach to assess an individual's unconscious aggressiveness. In Korea, psychological assessments that measure unconscious levels of aggression are lacking. Although the Korean CRT-A does not seem to assess implicit aggression in the same way that it does among Americans, the idea of CRT-A assessing an individual's aggressiveness is intriguing. Developing a Korean version of the CRT-A using the same idea (i.e. assessing unconscious aggressiveness using inductive reasoning) and having premises based on Korean culture would provide valuable information in understanding Koreans' aggressiveness and cross-cultural similarities and differences in individuals' aggressiveness.

In addition, this study provides meaningful information to understand faking issues regarding the personality assessments across cultures. Response distortion on the self-reported measures is prevalent (Amelang, Schäfer, & Yousfi, 2002; Piedmont, McCrae, Riemann, & Angleitner, 2000; Rogers, Sewell, Martin, & Vitacco, 2003; Viswesvaran & Ones, 1999) and respondents can fake their responses if they are motivated to do so. This could be due to the transparency of items on the self-reported personality survey. Even if participants are not told that the measure is a personality survey they can easily find out the purpose of the survey if they read the items on the self-reported personality measures. Conversely, people cannot see through the purpose of each item on the CRT-A, nor do they know that the CRT-A is a personality survey. Thus,

they cannot distort their responses on the CRT-A. This study demonstrates that the CRT-A remains resistant to faking, even among the Korean population.

Limitations

As limitations of this study, sample, criterion-related validity, between subject design for faking issues, and translation need to be discussed. First, the sample sizes of both groups were relatively small to conduct factor analyses. Although minimum sample size required for factor analysis is inconsistent, ranging from 50 to 1,000 (Arrindell & van der Ende, 1985; Comrey & Lee, 1992; Gorsuch, 1983; Kline, 1994), the larger the samples size, the better (Cudeck & O'Dell, 1994; MacCallum, Widaman, Zhang, & Hong, 1999). An especially large (> 1,000) sample size with CRT-A is recommended because the base rate of aggressive alternatives is low. As expected, aggressive individuals are rare and because of a low base rate of selecting aggressive alternatives, some items show zero to very small variance.

Second, this study investigated cross-cultural issues at measurement levels without testing their predictive validity in the Korean population. To be a valid measure of different cultures, construct validity should be tested; testing its criterion related predictive validity would provide additional meaningful information. For instance, as previously mentioned, the CRT-A tends to predict Americans aggressiveness (i.e., sabotage, lying, absenteeism, stealing, obstructionism, etc.); however, obtaining such hard criteria related to the Korean CRT-A was not practically possible. Therefore, predictive validity of the CRT-A has been not tested.

Third, for the translation process, this study only adapted one traditional method, back-translation, although it was additionally reviewed by a Korean professor. The Korean CRT-A did not show many errors from translation based on the back-translation process, and any discrepancies were resolved before conducting the study. However, the results suggest that some items were vague and unclear to Korean respondents. A little finesses and choice of word seems to affect participants' response patterns.

Future Directions

Implicit personality assessment, and understanding unconscious levels of personality through assessments, is new and fascinating to Koreans and research areas in cross-cultural studies of implicit personality assessments are fruitful subjects. First, Americans' responses supported that AQ and the CRT-A are less likely to be correlated and they predicted different types of aggression (i.e., verbal hostility, physical aggression, obstructionism). Previous research suggests that understanding aggression using both explicit and implicit measurements provides much more meaningful information than does using one or the other in predicting individuals' behaviors. Thus, it would be interesting to investigate the relationship between the self-reported AQ and the CRT-A in relation with criteria and the association between the two assessments among Koreans.

Second, a modified version of the CRT-A based on Korean culture would fit better with the Korean population and understanding their unconscious motives to be aggressive. The results of this study suggest that there were some CRT-A problems that may not be familiar to Korean culture; therefore, it would be intriguing to modify the CRT-A to align it more closely with Korean culture. The idea of the CRT-A assessing an

individual's unconscious motives through an inductive reasoning problem is fascinating and will truly provide valuable information in understanding Koreans' unconscious level of aggressiveness. Premises that are more familiar to Korean culture will more accurately assess their implicit aggression.

Third, this study failed to replicate the four-factor structure of the AQ. As mentioned above, respondents from Hong Kong China also did not support the four-factor structure, but they showed a good fit with a 12-item model of the AQ. Therefore, for future study, exploring measurement invariance with a shorter version of the AQ and CRT-A (if possible) might produce different results. Furthermore, completing short versions of the AQ and the CRT-A will take less time than completing the full versions, thus making easier to recruit more participants, leading to a larger sample, which will create a more concrete factor structure.

REFERENCES

- Amelang, M., Schäfer, A., & Yousfi, S. (2002). Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychologische Beiträge*, 44, 24-42.
- Anderson, E. (1994). The code of the streets. *The Atlantic Monthly*, 81-94.
- Ang, R. P. (2007). Factor structure of the 12-item aggression questionnaire: Further evidence from Asian adolescent samples. *Journal of Adolescence*, 30, 671-685.
- Archer, J., Holloway, R., & McLoughlin, K. (1995). Self-reported physical aggression among young men. *Aggressive Behavior*, 21, 325-342.
- Archer, J. & Webb, I. A. (2006). The Relation Between Scores on the Buss–Perry Aggression Questionnaire and Aggressive Acts, Impulsiveness, Competitiveness, Dominance, and Sexual Jealousy. *Aggressive Behavior*, 32, 464-473.
- Arrindell, W. A. & van der Endem J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9, 165-178.
- Aupperle, K. E. (1984). An empirical measure of corporate social orientation. *Research in Corporate Social Performance and Policy*, 6, 627-654.
- Auton, H. R., Pope, J., & Seeger, G. (2003). It isn't that strange: Paranormal belief and personality traits. *Social Behavior and Personality*, 31, 711-720.
- Aycicegi, A., Dinn, W. M., & Harris, C. L. (2005). Validation of Turkish and English Versions of the Schizotypal Personality Questionnaire-B. *European Journal of Psychological Assessment*, 21, 34-43.
- Barger, B. Nabi, R. Hong, L. Y. (2010). Standard back-translation procedures may not capture proper emotion concepts: A case study of Chinese disgust terms. *Emotion*, 10, 703-711.
- Bergeron, N. & Schneider, B. H. (2005). Explaining cross-national differences in peer-directed aggression: A quantitative synthesis. *Aggressive Behavior*, 31, 116-137.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92, 410-424.

- Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D. & James, L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology, 92*, 722-722.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Brenner, O. C. & Tomkiewicz, J. (1980). Relationship between aggression and managerial effectiveness. *Psychological Reports, 47*, 271-274.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.
- Bryant, F. B. & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss–Perry Aggression Questionnaire. *Journal of Research in Personality, 35*, 138-167.
- Burroughs, S. M. (2001). *The role of dispositional aggressiveness and organizational injustice on deviant workplace behavior* Unpublished manuscript, University of Tennessee, Knoxville.
- Buss, A. H. & Perry, M. (1992). The Aggression Questionnaire. *Journal of personality and Social Psychology, 63*, 452-459.
- Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155-175.
- 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.-S., 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, G. W. & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Research, 31*, 187-212.
- Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality, 69*, 979-1006.
- Clark III, I. (2000). Extreme response style in cross-cultural research: An Empirical investigation. *Journal of Social Behavior and Personality, 15*, 137-152.

- Comrey, A. L. & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, M. (1993). *Personnel selection and productivity*. New York, NY: Wiley.
- Costa, P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Crede, M., Bashshur, M. & Niehorster, S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, 92, 390-300.
- Cudeck, R. & O'Dell, L. L. (1994). Applications of standard error estimation in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, 115, 475-487.
- Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood Cases. *Psychological Bulletin*, 101, 126-135.
- Dodge, K. A. & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development*, 53, 620-635.
- Dodge, K. A. & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology*, 53, 1146-1158.
- du Toit M. (2003). IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESFACT, Scientific Software International.
- Dumka, L. E., Stoerzinger, H. D., Jackson, K. M., & Roosa, M. W. (1996). Examination of the cross-cultural and cross-language equivalence of the parenting self-agency measure. *Family Relations*, 45, 216-222.
- Embretson, S. E. & Reise, S. T. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Fink, M. (1963). Cross validation of an underachievement scale. *California Journal of Educational Research*, 14, 147-152.

- Fossati, A., Maffei, C., Acquarini, E., & Di Beglie, A. (2003). Multigroup confirmatory component and factor analyses of the Italian version of the Aggression Questionnaire. *European Journal of Psychological Assessment, 19*, 54-65.
- Frost, B. C., Ko, C-H. E., & James, L. R. (2007). Implicit and explicit personality: A test of a channeling hypothesis for aggressive behavior. *Journal of Applied Psychology, 92*, 1299-1319.
- Galim, A. & Avraham, S. (2009). A cross-cultural perspective on aggressiveness in the workplace: A comparison between Jews and Arabs in Israel. *Cross-Cultural Research: The Journal of Comparative Social Science, 43*, 30-45.
- Gay, P. (1993). *The cultivation of hatred*. New York, NY: W. W. Norton.
- Gentile, D. A., Lynch, P. J., Linder, J. R., & Walsh, D. A. (2004). The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. *Journal of Adolescence, 27*, 5-22.
- Gerevich, J., Bácskai, E., & Czobor, P. (2007) The generalizability of the Buss-Perry Aggression Questionnaire. *International Journal of Methods in Psychiatric Research, 16*, 124-136.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum.
- Greendlinger, V. & Byrne, D. (1987). Coercive sexual fantasies of colleges men as predictors of self-reported likelihood to rape and overt sexual aggression. *Journal of Sex Research, 23*, 1-11.
- Gülgöz, S. (2002). Five-Factor Model and NEO-PI-R in Turkey. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 175–196). NewYork, NY: Kluwer Academic Publisher.
- Heine, S. J., Buchtel, E. B., & Norenzayan, A (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19*, 309-313.
- Hein, S. J., Lenhman, D. R., Peng, K. & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality & Social Psychology, 82*, 903-918.

- Hogan, J. & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology*, 74, 273-279.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions. *American Psychologist*, 51, 469-477.
- Huesmann, L. R. & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, 72, 408-419.
- Innamorati, M., Pompili, M., Ferrari, V., Cavedon, G., Soccorsi, R., Mancinelli, I., Lelli, M. & Tatarelli, R. (2006). Relationship between Type-A behavior pattern, aggression, and suicide in Italian university students. *Individual Differences Research*, 4, 185-193.
- Ireland, J. L. & Archer, J. (2004). Association between measures of aggression and bullying among juvenile and young offenders. *Aggressive Behavior*, 30, 29-42.
- Jackson, D. N. (1984). *Personality research form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.
- James, L.R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, 1, 131-163.
- James, L. R. & LeBreton, J. M. (2011). *Assessing the implicit personality through conditional reasoning*. Washington, DC: American Psychological Association.
- James, L. R., McIntyre, M. D., Glisson, C. A., Bowler, J. L., & Mitchell, T. R. (2004). The conditional reasoning measurement system for aggression: An overview. *Human Performance*, 17, 271-295.
- James, L. R., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., et al. (2005). Conditional reasoning: An efficient, indirect method for assessing implicit cognitive readiness to aggress. *Organizational Research Methods*, 8, 69-99.
- James, L. R., & Mazerolle, M. D. (2001). *Personality in work organizations*, Thousand Oaks, CA: Sage.
- Jöreskog, K. G. & Sörbom, D. G. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software.
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction. *Journal of Applied Psychology*, 87, 530-541.

- Judge, T. A., Higgins, C.A., Thoresen, C.J., & Barrick, M.R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621–652.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87, 797-807.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kelly, J. A. & Worell, L. (1978). Personality characteristics, parent behaviors, and sex of subject in relation to cheating. *Journal of Research in Personality*, 12, 179-188.
- Kim, S., Kim, S-H, & Kamphaus, R. W. (2010). Is aggression the same for boys and girls? Assessing measurement invariance with confirmatory factor analysis and item response theory. *School Psychology Quarterly*, 25, 45-61.
- Kline, P. (1994). *An easy guide to factor analysis*. New York, NY: Routledge.
- Knight, G. P. (2000). Measurement issues in “Assessing the home environments of young adolescents:” A commentary, *Journal of Research on Adolescence*, 10, 299-305.
- Ko, C. E., Thompson, V. M., Shim, H. S., Roberts, J. S., & McIntyre, H. (2008). Alternative scoring strategies for the conditional reasoning test of aggression. In J. LeBreton (Chair), *Psychometric properties of conditional reasoning tests*. Symposium conducted at the 23rd Annual Conference of the Society of Industrial and Organizational Psychology, San Francisco, CA.
- LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, 92, 1-16.
- Lee, C. & Green, R. T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies*, 22, 289-305.
- Lee, H. J. (2009). *CRT-RMs cross-cultural study with Korean college students*. Unpublished master’s thesis, Georgia Institute of Technology, Atlanta.
- Lee,^a S. (2001). *Cross-cultural validity of personality traits for predicting job performance of Korean engineers*. Unpublished doctoral dissertation, Ohio State University, Columbus.
- Lee,^b J. (2001). *Relationship between personality traits and English proficiency among Korean university students*. Unpublished doctoral dissertation, University of Mississippi.

- Lodhi, P. H., Deo, S., & Belhekar, V. M. (2002). The Five-Factor model of personality in Indian context: measurement and correlates. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 227–248). New York, NY: Kluwer Academic/Plenum Publishers.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Markus, H. R. & Kitayama, S. (1991) Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224-253.
- Maxwell, J. P. (2007). Development and preliminary validation of a Chinese version of the Buss-Perry Aggression Questionnaire in a population of Hong Kong Chinese. *Journal of Personality Assessment*, 88, 284-294.
- McCrae, R.R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R.R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 105–126). New York, NY: Kluwer Academic/Plenum Publishers.
- McCrae, R. R. & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509-516.
- McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36, 587–596.
- McCrae, R. R., Costa, P. T., Pilar, G. H., Rolland, J. P., & Parker, W. D. (1998). Cross-cultural assessment of the five-factor model: The revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 29, 171–188.
- McLoyd, V. C. (1999). Cultural influences in a multicultural society: Conceptual and methodological issues. In A. S. Masten (Ed.), *The Minnesota Symposium on Child Psychology: Vol. 29. Cultural processes in child development* (pp. 123-135). Mahwah, NJ: Erlbaum.
- Miotto, P., De Coppi, M., Frezza, M., Petretto, D. R., Masala, C., & Preti, A. (2003). Eating disorders and aggressiveness among adolescents. *Acta Psychiatrica Scandinavica*, 108, 183-189.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683-729.
- Motowidlo, S. J., Hooper, A. C., Jackson, H. L. (2006). Implicit Policies About Relations Between Personality Traits and Behavioral Effectiveness in Situational Judgment Items. *Journal of Applied Psychology*. 91, 749-761.

- Murray, H. A. (1938). *Explorations in personality*. New York, NY: Oxford University Press.
- Muthén, L. K. & Muthén, B. O. (2006). *Mplus 4.2. [Computer software]*. Los Angeles Author.
- Nakano, K. (2001). Psychometric evaluation on the Japanese adaptation of the Aggression Questionnaire. *Behavior Research and Therapy*, 39, 853-858.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias. *European Journal of Social Psychology*, 15, 263-280.
- Noh, S., Avison, W. R. & Kaspar, V. (1992). Depressive symptoms among Korean immigrants: Assessment of a translation of the Center for Epidemiologic Studies—Depression Scale. *Psychological Assessment*, 4, 84-91.
- Patton, T. W. (1999). *Measuring personal reliability via conditional reasoning: identifying people who will work reliably*. Unpublished doctoral dissertation, University of Tennessee, Knoxville.
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values Across cultures and possible solutions. *Psychological Methods*, 2, 329-344.
- Piedmont, R. L. & Chae, J. H. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO-PI-R for Koreans. *Journal of Cross-Cultural Psychology*, 28, 131–155.
- Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T. Jr. (2002). The applicability of Five Factor Model in sub Saharan culture: The NEO-PI-R in Shona. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 155–174). New York, NY: Kluwer Academic/Plenum Publishers.
- Poortinga, Y. H., van de Vijver, F. J. R., & van Hemert, D. (2002). A Cross-cultural equivalence of the Big Five: A tentative interpretation of the evidence. In R.R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 105–126). New York, NY: Kluwer Academic/Plenum Publishers.
- Roesch, S. C., Wee, C., & Vaughn, A. A. (2006). Relations between the Big Five personality traits and dispositional coping in Korean Americans: Acculturation as a moderating factor. *International Journal of Psychology*, 41, 85-96.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.

- Russell, S. M. & James, L. R. (2008). Recording lying, cheating, and defiance in an internet based simulated environment. *Computers in Human Behavior*, 24, 2014-2025.
- Sablynski, C. J., & Mitchell, T. R. (2006, May). Post hoc analyses of interactions between the Aggression Questionnaire and Conditional Reasoning Test of Aggression. In J. M. Lebreton (Chair), *Bringing the Implicit Personality into I-O Psychology*. Symposium presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Sinaiko, H. (1963). Men, Machines and Systems. *PsycCRITIQUES*, 8, 297-298.
- Schmitt, D. P. & Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. *Journal of Personality and Social Psychology*, 89, 623-642.
- Stevenson, H. W., Lee, S., Chen, C., Stigler, J. W., Hsu, C., & Kitamura, S. (1990). *Contexts of achievement: A study of American, Chinese, and Japanese children. Monographs of the Society for Research in Child Development*, 55 (Serial No. 221).
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.) *Differential Item Functioning* (pp. 67-113). Mahwah, NJ: Lawrence Erlbaum.
- Thissen, D., Sterinberg, L., & Wainer, H. (1998). Use of item response theory in the study of group differences in trace lines. In H. Wainer & I. Brum (Eds.) *Test Validity* (pp. 147-169). Mahwah, NJ: Lawrence Erlbaum.
- Trapmann, S., Hell, B., Hirn, J.-O.W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Journal of Psychology*, 215, 132-151.
- Toch, H. (1993). Good violence and bad violence: Through self-presentations of aggressors accounts and war stories. In R.B. Felson & J. T. Tedeschi (Eds.) *Aggression and violence: Social interactionist perspectives* (pp. 193-206). Washington, DC: American Psychological Association.
- Tsutsumi, A., Iwata, N., Watanabe, N., de Jonge, J., Pikhart, H., Fernández-López, J. A., Xu, L., Peter, R., Knutsson, A., Niedhammer, I., Kawakami, N., & Siegrist, J. (2009). Application of item response theory to achieve cross-cultural comparability of occupational stress measurement. *International Journal of Methods in Psychiatric Research*, 18, 58-67.

- van de Vijver, F. J. R. & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.
- van de Vijver, F. J. R. & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality*, 69, 1007-1032.
- Viswesvaran, C. & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational Psychological Measurement*, 59, 197-210.
- Werner, O. & Campbell, D.T. (1970). Translating, working through interpreters and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398-420). New York, NY: The Natural History Press.
- Wheeler, M. S. & Acheson, S. K.(1993). Criterion-related validity of the Life-Style Personality Inventory. *Individual Psychology: Journal of Adlerian Theory, Research & Practice*, 49, 51-57.
- Woo, J. M. (2009). *Psychology of Men*. Seoul, Korea: ReadersBook.
- Wu, C.-H. (2008). An examination of the wording effect in the Rosenberg Self-Esteem scale among culturally Chinese people. *The Journal of Social Psychology*, 148, 535-551.