

# Psychological Well-being Measurement: A Comparative Study of Korean and American Adults

Jeong-shin An\*, Michael C. Lambert\*\*, Gyoung-hae Han\*\*\*, Seung-eun Cha\*\*\*\*

Researcher, Center for Family Policy & Research, University of Missouri-Columbia\*.

Professor, Dept. of Human Development and Family Studies, University of Missouri-Columbia\*\*

Professor, Dept. of Child Development and Family Studies, Seoul National Univ.\*\*\*

Ph.D Candidate, Dept. of Child Development and Family Studies, Seoul National Univ.\*\*\*\*

**Abstract :** Ryff's (1989) psychological well-being measure is used to assess and sometimes compare Korean and American adults, however, there is no information regarding whether its dimensions are psychometrically invariant across, whether its items provide sufficient information for, and whether each item measures identical trait levels in, the two nations. Confirmatory factor analysis on response 1,696 Korean and 3,669 American adults, gave to the measure revealed lack of fit and absence of factorial invariance across the two nations. Item response theory revealed significant variance for items on each factor across two countries that most items yielded limited psychometric information. And that each item measure different trait levels, suggesting that in its present form, the measure might lead to misleading results for, and across the two nations.

**Key Words :** psychological well-being, cultural differences, confirmatory factor analysis, item response theory

## I. Introduction

Most existing studies of adult psychological well-being have focused on physical health or the lack of mental health problems (Keyes & Ryff, 1998). However, psychological well-being is not the opposite of dysfunction or lack of illness. In other words, positive psychological functioning is not less unhappiness or suffering. Ryff (1995) has emphasized the need to consider psychological well-being as a multidimensional positive construct. By integrating established life span development theories (e.g., Erikson, Maslow,

Jung, Neugarten) and from clinical descriptions of positive mental health, she proposed a six dimensional model for psychological well-being, which she labeled as autonomy, self-acceptance environmental mastery, personal growth, purpose in life, and positive relation with others (Ryff, 1989; Ryff & Heidrich, 1997). First, Ryff and her colleagues developed a multidimensional measure consisting of 120 items to measure her theoretical dimensions of psychological well-being. Additionally, Ryff and Keyes (1995) tested a shortened version consisting of 18 items of her psychological well-being measure with

---

Corresponding Author: Jeong-shin An, Center for Family Policy & Research, University of Missouri-Columbia, Korea  
E-mail: AnJS@missouri.edu

confirmatory factor analysis and the results confirmed the six dimensional model.

Although Ryff purports that the measure provides psychometrically sound information limited to American adults, it is widely used to assess adults in other nations and also to compare levels of psychological well-being in adults cross-nationally (Clarke, Marshall, Ryff, & Rosenthal, 2000; Clarke, Marshall, Ryff, & Wheaton, 2001; Kim & Kim, 2000; Ruini, Rafanelli, Ryff, & Fava, 2003; Ryff, Lee, & Na, 1993). It is important to demonstrate that all measures used for the assessment of multiple groups possess adequate estimated psychometric properties for each group. Additionally, when such measures are used to compare the functioning of different groups it is important to demonstrate that its factorial model and items are invariant across such groups (Diner, Oishi, & Lucas, 2003; Lambert et al., 2003; Lu, Gilmour, & Kao, 2001; Romero, 2000; Van de Viver, 2003). Without information on the psychometrical soundness of comparing findings, regarding the functioning of individuals from different cultural backgrounds, it is difficult to determine whether findings from such comparisons reflect true differences or similarities between the two groups or whether such findings might be artifacts of measurement. For example, Ryff (1995) used her psychological well-being measure to study both U.S. and South Korean adults. She reported that persons from South Korea reported higher levels of positive relations with others and lower levels of self-acceptance than their U.S. counterparts. Additionally American adults reported high levels of personal growth and

purpose in life. In interpretation of these findings, Ryff (1995) inferred that cultural variation in self-presentation might explain some of the psychological well-being differences for adult reports across the two nations.

Despite its use in Korea and the United States to compare the functioning of adults across both nations, there is insufficient evidence regarding the psychometric soundness of using the psychological well-being measure to assess adults within and across the two nations. For example, we have no information, pertaining to the existence of measurement invariance across Korea and the United States. Invariance implies that target individuals with similar levels of psychological well-being rate their functioning on an identical metric (Cooke, Kosson, & Michie, 2001). Lack of invariance can have serious consequences for assessment and research, as cross-national differences observed in informants' reports could be the reflection of measurement bias across the nations surveyed (Reise, Widaman, & Pugh, 1993). Thus, findings from cross-national studies could be attributed whether it is the measurement bias, or true group differences, or is a combination of both situations (Cooke et al., 2001; Millsap, 1997; Reise et al., 1993). The implicit assumption that informants across the two nations provide psychometrically invariant information on their psychological well-being cross-nationally could be correct. Nevertheless, this assumption remains untested. Thus, the main purpose of this study is to test factorial and item invariance of Ryff's psychological well-being measurement for Korean and U. S. adults.

## **II. Methods Used to Assess Invariance**

Although other procedures such as the Mantel-Haenszel (Holland & Thayer, 1988) or logistic regression (Swaminathan & Rogers, 1990) are used to examine invariance, it is most commonly tested by two methods: confirmatory factor analysis (CFA) and item response theory analyses (IRT) (Raju, Laffitte, & Byrne, 2002; Reise et al., 1993). As we stated above, CFA studies have been conducted on Ryff's psychological well-being model (Clarke, Marshall, Ryff, & Wheaton, 2001; Ryff & Keyes, 1995), but research examining factorial invariance across any two or more groups appear to be nonexistent. We also found no IRT-based studies on the Psychological Well-being measure.

### **1. Confirmatory Factor Analysis**

Factor analysis is classified as exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis the characteristics and relationships among items (variables) in a database and thus provides factors that summarize such relationships. By contrast, confirmatory factor analysis uses covariance matrices to compare a theorized model within a database. An important strength of CFA is its ability to test for model fit and factorial invariance of multidimensional models across two or more groups. Nevertheless, CFA possesses certain drawbacks. Similar to exploratory factor analyses (EFAs), its results are usually sample-dependent and not always generalizable to other groups (Embretson & Reise,

2000; Mellenbergh, 1996). Moreover, once factor models are derived or confirmed, measurement designers and users usually give equal weightings to all items derived from each factor by routinely totaling scores on items loading on a factor of interest. This practice does not account for the trait level (in the case of the psychological well-being) that each item loading on any specific factor estimates. In effect, as usually employed, CFA does not assess invariance at the item level.

### **2. Item Response Theory and Differential Item Functioning**

Methodology subsumed under the broad umbrella of classical test theory-based methodology (CTT) is routinely used for development and validation of most psychological measures (Embretson & Reise, 2000). Estimates of trait levels are typically obtained from summing of the scores from responses to items of a specific dimension while ignoring trait levels such items measure. By contrast item response theory (IRT) according the pattern of responses and the trait levels reflected in such patterns (Cooke & Michie, 1997). Theoretically, IRT methodology is item independent allowing the administration of different subsets of psychometrically sound items from a dimension of focus while obtaining valid information regarding the trait on such item subsets (Mellenbergh, 1996; Reeve, 2003; Weiss, 1995). IRT also provides psychometric information for each item on a trait of focus and the trait level at which the best discrimination and the least amount of measurement error occur (De Ayala, 1993; Mellenbergh, 1996). Moreover, it can

determine whether identical amounts of psychometric information are given at identical trait levels across different groups such as the ratings given by Korean and American adults.

In IRT models, latent traits are usually represented by the symbol  $\Theta$ . Theoretically,  $\Theta$  values range from negative to positive infinity. For psychometric measures where it is presented graphically in standardized score format on the X-axis,  $\Theta$  ranges from -3 to +3 (Embretson & Reise, 2000). The probability of respondents choosing to item response options that matches their trait levels is presented on the Y-axis. The parameter  $b_{jk}$  locates the scale value of the response options for an item on the X-axis, where higher  $b_{jk}$  parameters are associated with higher  $\Theta$  levels. The discrimination parameter,  $a_{jk}$ , is usually between the values of 0 and 3, with larger values associated with steeper slopes, better item discrimination between different  $\Theta$  levels, and more precise information on the trait levels the item measures (Gray-Little, Williams, & Hancock, 1997).

The relationship between the  $a_{jk}$  and  $b_{jk}$  parameters is displayed on the same graph as  $\Theta$  using logistic curves, known as operating characteristic curves (OCCs) (Embretson & Reise, 2000). The number of response categories per item determines the number of  $b_{jk}$  parameters and the number of  $b_{jk}$  parameters is  $k-1$ , where  $k$  is the number of response categories (Embretson & Reise, 2000). The well-being form that Keyes and Ryff developed has seven response categories. Therefore, in its original form, each well-being item has six  $b_{jk}$  parameters. The Korean version of the psychological well-being measure uses a five-point Likert scale. Thus, each well-being

dimension has four  $b_{jk}$  parameters. It is important to note that IRT models provide error estimates for each estimated item parameter.

IRT models can be used in development and validation of both dichotomous rating scales and polytomous rating scales. Samejima's (1969) graded response model is one of the appropriate IRT models for analyses of the Likert scale data such as those of the present study. The application of Samejima's model is more fully discussed in the Methods and Results. IRT also provides the professional with estimates regarding the amount of information an item provides and the  $\Theta$  level where the item provides the most information are. Such item information is usually plotted on item information curves (IIC).

For items with polytomous rating scales, the probability of an informant responding to a particular category is calculated from the  $b_{jk}$  parameters and they are displayed graphically with item trace lines called item characteristic curves (ICC). Therefore, the item characteristic curves are derived from the  $a$  and  $b$  parameters for each item on each well-being dimension and  $\Theta$  levels are also displayed on the x-axis while the probability of a given response for that item category is displayed on the y-axis. The ICCs represent the probability of responding affirmatively to a category given the respondent's trait level. Because psychological well-being items are rated on a 1 through 7 rating scale, seven trace lines are calculated and plotted for each item. Since the Korean items are rated on a five-point Likert scale. There are five trace lines reflecting each point on the Likert scale. In addition to ICC, IRT provides information for the entire dimension of interest, which is plotted on the test

information curve (TIC).

IRT has the power to reveal items that are invariant. That is, items with nonsignificant differential item functioning (DIF) versus those that lack this quality. Broadly defined, DIF indicates that a particular item has different OCCs for different groups of people, so that individuals from one of these groups and individuals from the other group with identical trait levels have a different probability of choosing a particular option for the item of focus (Robie, Zickar, & Schmit, 2001). Applied to the well-being measure, DIF would occur if adults with identical levels of functioning across two or more nations and who complete the Psychological Well-being measure do not have the same expected probability of endorsing the same response category on the item of interest (Park & Lautenschlager, 1990). Cooke and Michie (1999) used the example of Celsius versus Fahrenheit scales as a striking example of DIF. Both scales measure temperature, but they both have different zero points and the intervals between units of measurement on each scale differ. Comparing ratings given by two or more sets of respondents that lack measurement invariance would be as meaningless as comparing measurements of temperature from both Celsius and Fahrenheit scales without adjustment for their scalar differences.

Consensus regarding disposition of items with DIF is nonexistent (Zickar & Drasgow, 1996). One option that the professional has, is discarding items with DIF while retaining those that have nonsignificant DIF, especially those that furnish adequate amounts of information across groups. This decision could permit professionals to validly measure the construct of interest across focus

groups. Such a procedure could be psychometrically tenuous if one exclusively uses CTT-based procedures.

Besides invariance across responses given by adults from different nations, IRT also has the ability to determine whether each item within a dimension for a particular group of informants measures the same trait levels. Most professionals routinely total respondents' ratings to all items on dimension, a practice which could result in the possibility that on a given dimension an adult with trait levels reflecting limited psychological well-being could receive the same or even a higher score than one with considerably higher levels of well-being. Using IRT scoring procedures that examine different patterns of responses could better elucidate the differences in levels of well-being across the two nations. In IRT analyses, items that represent high levels of psychological well-being could discriminate well and be weighted highly in an IRT dimension score only when there is other evidence that an examinee is experiencing high levels of psychological well-being. If the pattern of responses to other items suggests low levels of well-being, then these items would receive little or no weight in the estimation of their psychological well-being.

### **3. Purpose of Study**

Focusing on psychological well-being, the present study is designed to address the following objectives. First, we will test for unidimensionality in each psychological well-being factor. Second, we will examine whether factorial invariance exists for Ryff's psychological

well-being model across the two nations. Third, we will try to show whether the parameters for each item on each psychological well-being dimension are invariant across ratings provided by Korean and American adults. Fourth, we will look for item parameter estimates for each item on each dimension for Korean and American adults. Finally, by examining the item information curve for each item, we would be able to provide guidelines regarding which items might be retained or removed from each dimension.

### III. Method

#### 1. Sample

The U.S. data were obtained from the National Survey of Midlife Development in the United States (MIDUS), a national study conducted in 1995 by The John D. and Catherine T. McArthur Foundation Network, which focused on the factors contributing to successful midlife development

(MIDMAC) (Brim et al., 2000). The sample of 4,242 adults, ages 25 to 74, responded to an initial telephone interview and then a follow-up mail questionnaire. The response rate was 70% for the telephone interview and 87% for the follow-up mail questionnaire of telephone respondents. For the present study, cases that had missing values for all psychological well-being items were removed from the analyses. The final sample used in present study included 3,710 adults.

Korean data were derived from a study designed to gather information on Study for Successful Development of Midlife in Korea (Han et al., 2001). Trained interviewers visited each respondent's home and collected data by face-to-face interview on 1,696 Korean adults, ages 27 to 60 from 16 states. <Table 1> shows that samples from both nations consisted of approximately equal numbers of males and female.

#### 2. Measure

To assess adults' psychological well-being in both

<Table 1> Characteristics of Korean and American Adult Samples

Variables		Korean (N=1,696)	U. S. adults (N=3,710)
Gender	Male	857(50.5%)	1,850(49.9%)
	Female	839(49.5%)	1,860(50.1%)
Age	20s	30( 1.8%)	310( 9.3%)
	30s	684(40.3%)	771(23.0%)
	40s	523(30.8%)	834(25.0%)
	50s	455(26.8%)	692(20.7%)
	60s	4( 0.2%)	518(15.5%)
	70s		218( 6.5%)
Educational Level	Lower than High school	195(17.4%)	530(15.7%)
	Graduate High school	826(48.7%)	1203(40.3%)
	Higher High school	572(33.7%)	1637(45.0%)

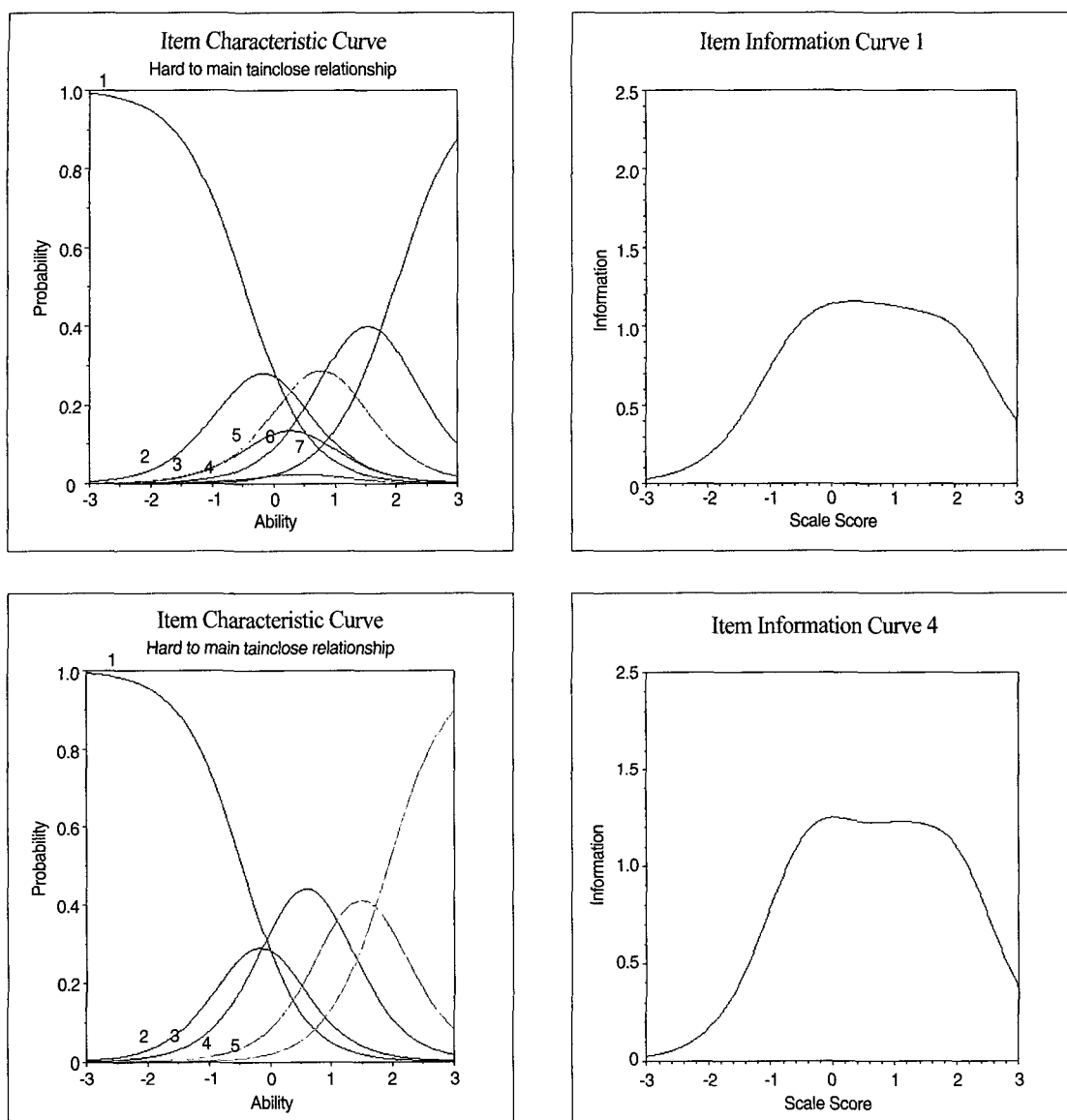
nations, the 18-item short-form Psychological Well-being scale developed by Ryff (1989) was used. As outlined in the introduction and briefly summarized here, Ryff's psychological well-being model consists of six dimensions, with items that measure psychological challenges toward Autonomy, Environmental Mastery, Personal Growth, Positive Relations with Others, Purpose in Life, and Self Acceptance (Keyes & Ryff, 1998; Ryff, 1989). According to CTT-based psychometric indices, each item from each dimension considered together as part of a unidimensional well-being scale is documented to possess adequate internal consistency for U.S. ( $\alpha=.80$ ) and for Korean ( $\alpha=.61$ ) adults.

Respondents from the United States rated each well-being item on a 7-point Likert-type scale ranging from "1 = *strongly agree*" to "7 = *strongly disagree*". Korean adults used 5-point Likert scale ranging from "1 = *strongly agree*" to "5 = *strongly disagree*". For invariance test across two nations, the 7-point Likert scale of U. S. adults group was re-categorized as 5-point scale after checking their Item Characteristic Curves. The curves for each item show that there was considerable overlap between categories 3, 4, and 5 for the U.S. sample and that the respondents made little discrimination between these points on the Likert scale. Thus, we merged all such categories into one category. Figure 1 provides an example for the item "Maintaining close relationships has been difficult and frustrating for me" an item under the Positive Relationships with Others dimension. It should be noted that <Figure 1> shows virtually no difference between the amount of information for the seven-point Likert scale item and the re-categorized with 5-point scale.

### 3. Data Analysis

All CFAs were performed using AMOS 4.0 (Arbuckle & Wothke, 1999) maximum likelihood estimation. The CFA methodology provides fit indices, used to assess whether a hypothesized model fits the researcher's data (DeShon, 1998). The  $\chi^2$  is the only true statistical test that assesses model fit, but it is sensitive to large sample sizes (Schumacker & Lomax, 1996) and can result in inappropriately rejecting a well-fitting model. Therefore, the Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) all of which are insensitive to sample sizes, were also used to judge model fit. Models with TLI and CFI  $> .90$  and RMSEA  $< .05$  are deemed to be of good fit. TLI and CFI indices between .8 and .9 and RMSEA values between .05 and .08 are considered to be moderate fit (Steiger, 1990). It should be noted that Hu and Bentler (1998) suggested slightly higher values for the CFI and TLI (i.e., .95) as indicators of good fit although more recently others (Vandenberg & Lance, 2000) indicated that it might be premature to accept such stringent fit indices.

Because the rating scales for the psychological well-being measurement are in Likert scale format, we applied Samejima's graded response model to the data (Samejima, 1969). For all such analyses we used the MULTILOG 7.0 program (Scientific Software International, Inc., 2002). MULTILOG provides statistical analyses where observed data are compared with a theorized model. This model comparison is evaluated by a -2 X likelihood (i.e.,  $G^2$ ) statistic with a distribution that approximates



<Figure 1> Item Characteristic Curves for the Item "Maintaining close relationships has been difficult and frustrating for me": 7-point scales and 5-point scales

the chi-square and its degrees of freedom based on the number of parameters estimated (Thissen, Steinberg, & Wainer, 1991).

To assess DIF, the MULTILOG application requires that the data be structured as if they are

separate items for each group. In the present study where responses given by two sets of informants were studied, each studied item is represented as two different items, with one item having response data for the reference group, but missing data for



the focal group. The converse is true for the data in the focal group, which were recoded as missing in the reference group. In the present study, we established Korean adults as the reference group and American adults as focal groups. Testing whether DIF exists across the groups requires the identification of one or more anchor items across the groups studied. Anchor items are DIF free and thus provide a common metric across a reference and one or more focal groups (Embretson, 2000; Reise, 1993).

First, we determined whether it was necessary to conduct this iterative procedure by performing a global test of whether one or more items in the dimension of focus possessed DIF (Reise et al., 1993). Thus, upon obtaining the  $G^2$  estimate from an unconstrained model for the syndrome of focus, we constrained the parameters for all items across groups. The fully constrained model was nested in the unconstrained model, where a nonsignificant  $\Delta G^2$  index would indicate an absence of DIF for one or more items and a significant  $\Delta G^2$  would indicate that DIF was evident in one or more items. When  $\Delta G^2$  indicated DIF presence for one or more items the next step was an iterative approach where the parameters for each item were constrained iteratively, and where the model with the one constrained item was nested in the unconstrained model. Items with nonsignificant  $\Delta G^2$  were considered as anchor items, while the others were considered as having item DIF and were candidates for study in the next step. We tested for DIF in each studied item by examining whether item parameter estimates and their resulting TICs had significantly different slopes across informants.

The IRT literature extensively discusses the

importance of item information yield in making decisions regarding test items that could be candidates for retention versus those that might be dropped. Unfortunately, our review of this literature revealed no specific guidelines regarding the amount of information that an item should yield to meet the retention or exclusion criteria. Therefore, we followed the two-step procedure. (1) For each dimension and the items that comprise them, we examined the test information curves (TIC) and the IIC of each item that are part of their respective dimension. (2) We noted the value of each  $a$  parameter and the position of the location parameters for each item on each dimension. That is, because the item information curve is a function of the  $a$  and  $b$  parameter estimates where the  $a$  parameter estimate determines the height of IIC and the  $b$  parameter estimate determines where such a curve peaks (Reise 1999). Thus, bearing in mind that the TIC is the sum of all IICs, we compared the contribution each IIC to its respective TIC. Besides noting that most TICs peaked well above 10, our observations revealed the following three patterns. 1) Items with a parameter estimates below 1 had virtually flat IICs with little semblance of peaks and their highest points were well below .1. 2) Items that had a parameter estimates slightly over 1 either yielded IIC peaks that could be slightly under .1 or even higher than .1 but that peaks were usually below .5. Finally we noted that items with a parameters well above 1 (e.g., > 1.5) often had IIC peaks that were above .5. IIC peaks that were less than or equal to .1, greater than .1 but less than .5, and greater than .5 were viewed as providing poor, mediocre, and good information respectively. Thus, items

meeting the first two criteria could be considered as candidates for removal while those in the third could be considered as candidates for retention.

## **IV. Results & Discussions**

### **1. Unidimensionality Test**

IRT analyses assume unidimensionality implying that for a given dimension, one latent factor underlies the data. To address the question of unidimensionality, we used exploratory factor analysis where the ratio of the first-to-second eigenvalues should be 3:1 (Bolt, Hare, Vitale, & Newman, 2004). The ratios of the first-to second eigenvalues are ranged from 11.2:1 (first eigenvalue is 2.71 and second one is 0.24) and 1.29:1 (first eigenvalue is 1.24 and second one is 0.96). Because the ratios did not closely approximate the unidimensionality criteria, we subject the data to exploratory factor analyses for both groups considered together and separately. Although we examined multiple factor solutions, all our tests for unidimensionality revealed ratios that were poorer than those derived for the factors of the original model. We therefore conducted all analyses on the existing dimensions.

### **2. Factorial Invariance Test**

We addressed the second objective of the study by using CFA procedures to provide estimates of model fit and to test whether factorial invariance existed for theoretical model for Ryff's psychological well-being measures for Korean and

American adults. Three types of theoretical models which Ryff and Keyes (1995) tested were examined in this study: a psychological multidimensional nonhierarchical model (i.e., a multidimensional nonhierarchical model with six sub-dimensions having each three indicators and where all first order factors covaried with one another), a psychological first order unidimensional hierarchical model (i.e., with the indicators loaded one psychological well-being factor), and second order model a with single second order. For our tests of structural invariance, we conducted analyses on two nested models for each of the three theoretical models. First we tested a baseline model (Reise et al., 1993) where fit of the covariance matrices for Korean and American adults were freely estimated. Second we tested a constrained model where all paths from latent variables to indicators across both Korean and American adults were constrained (Joreskog, 1971). A nonsignificant  $\Delta\chi^2$  would indicate factorial invariance across both nations.

<Table 2> presents the fit indices for all invariance tests conducted. The large chi-square value and all other fit indices indicated that tests conducted on variants of Ryff's (1989) models did not reveal good fit and that structural invariance did not exist for any of the three models. Nevertheless, the indices fit for Ryff's psychological well-being models approximated moderate fit. Therefore, we used a psychological multidimensional nonhierarchical model as our final model to examine IRT.

### **3. Item Invariance Tests**

Our third objective was to determine whether

&lt;Table 2&gt; CFA Invariance Tests for Korean and American Adults

Models	Ns	$\chi^2$ (df)	TLI	CFI	RMSEA	$\Delta\chi^2$
Variants of Existing Theoretical Psychological Well-Being Models						
Psychological Multidimensional Nonhierarchical	5406	8809.47(240)	.79	.83	.08	
Psychological Multidimensional Nonhierarchical-Constrained	5406	14249.96(258)	.68	.73	.10	5440.49*** (18)
Psychological First Order Unidimensional Hierarchical	5406	24366.17 (270)	.47	.53	.13	
Psychological First Order Unidimensional Hierarchical-Constrained	5406	38805.83 (287)	.20	.25	.15	14439.66*** (17)
Psychological One Second Order Well-Being Factor	5406	15871.11 (258)	.64	.70	.19	
Psychological One Second Order Well-Being Factor-Constrained	5406	16337.75 (263)	.63	.68	.10	466.64*** (5)

Note: The values in parentheses are degrees of freedom. TLI = Tucker Lewis Index, CFI = Comparative Fit Index and RMSEA = Root mean square error of approximation. All  $\Delta\chi^2$  values are calculated from the unconstrained model nested in the constrained models across Korean and U. S. adults.

invariance existed for item parameter estimates on Korean and American adults for dimensions from model that most closely fits the data as evident in CFA fit indices. <Table 3> indicates that across all of the dimensions studied only seven items provide invariant information across two nations.

The fourth objective of providing item parameter estimates is evident in <Table 3>(refer to Appendix) where all the item parameter estimates are listed. Indicating which items might be worthy of retention is also in <Table 3> where only items across all factors provide sufficient psychometric information to be retained for the assessment of Korean adults is nine items. Those for American adults are eleven items. Autonomy has two invariant items providing good information (“I have confidence in my opinion even they are different from the way most people think” and “I judge myself by what I think is important, not by the values

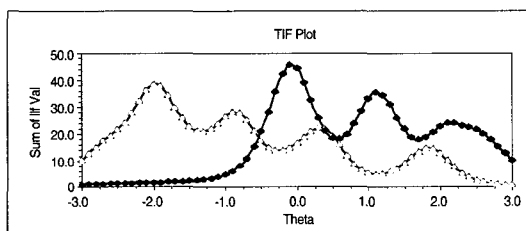
of what others think is important” across two countries. Two items provide invariant information for Self-Acceptance dimension. “I like most part of my personality” provides mediocre information and “When I look at the story of my life, I’m pleased with how things have turned out so far” provides good information for adults in both nations. Personal Growth and Environmental Mastery have each only one invariant item which provides good information for each dimension: “I think that it is important to have new experiences that challenge how I think about myself and the world” and “In general, I feel I am in charge of my life”. In Purpose in Life, there is no invariant item (see Table 3). For example, “I live life one day at a time and don’t think about the future” provides good information across two nations, but the parameters are different. Therefore, the item provides good information for both groups, but not the same information. Positive

Relations with Others has one good invariant information item, "I have not experienced many warm and trusting relationships with others". The other items provided different information across Korean and American adults.

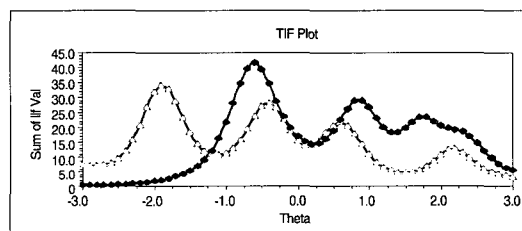
Since significant DIF emerged for items on dimension, the last objective was to examine the test information curves (TIC) and test characteristic curves (Figure 2 & 3, in page 14, 15, 16). Because item information functions for each item were aggregated to form the test information functions (TIC), each dimension's TICs were

compared for Korean and American Adult groups. <Figure 2> presents the result of TIC comparison between Korean and American Adult groups. For American adults, test information clearly increases at mid-and lower ranges of  $\Theta$ , and then gradually rises to the .0 and higher toward positive of each latent trait. In addition, each curves declines at the end of the positive  $\Theta$  scale. In case of Korean, test information curves scattered through the  $\Theta$  ranges and lower values than American adults. This finding indicates that all items on each dimension provide more information for adults in America

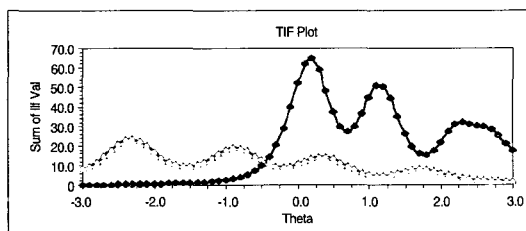
#### Autonomy



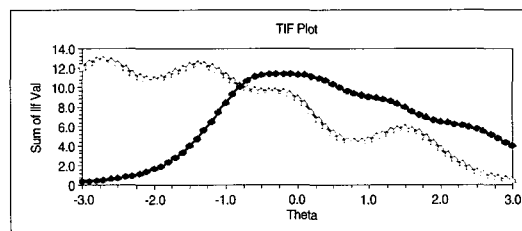
#### Self Acceptance



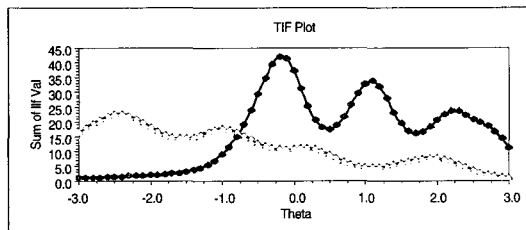
#### Personal Growth



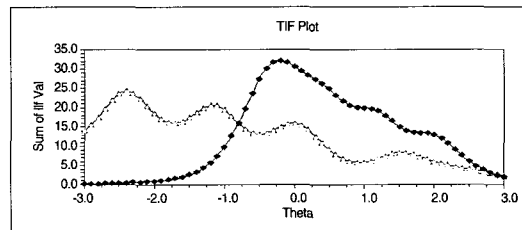
#### Purpose in Life



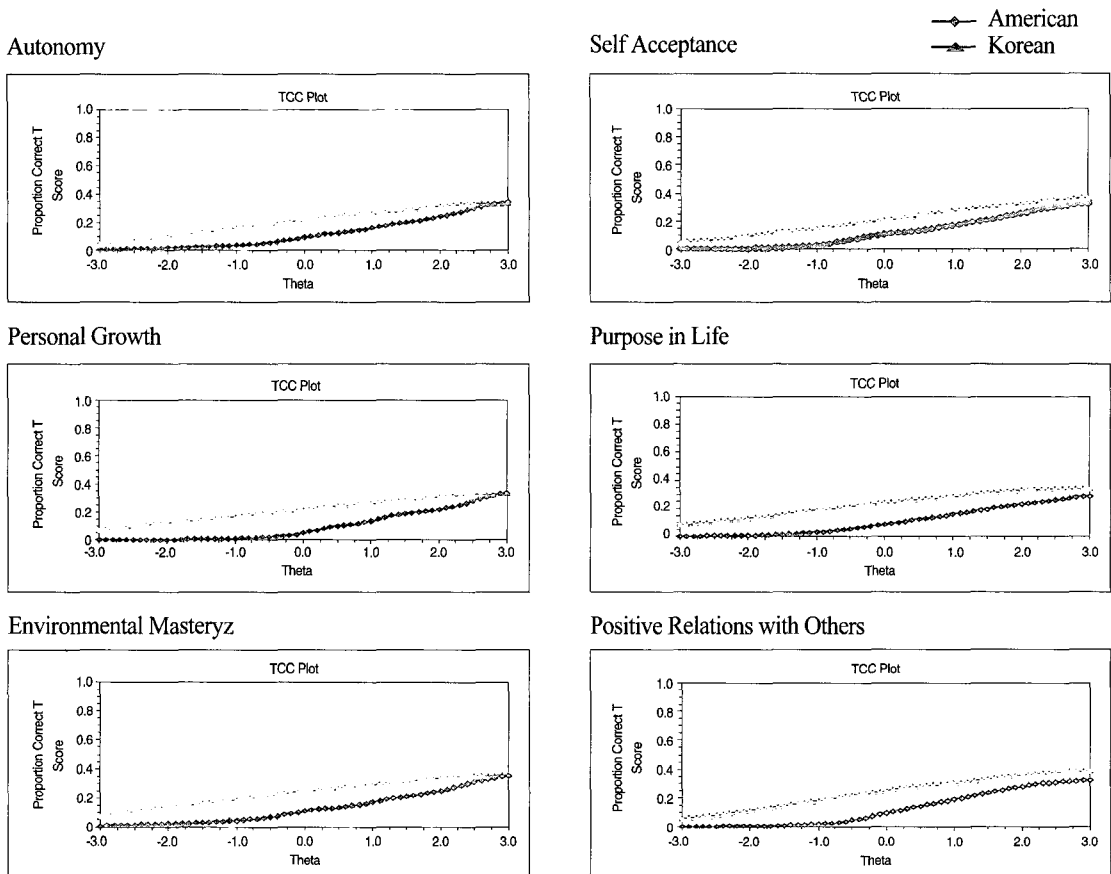
#### Environmental Mastery



#### Positive Relations with Others



<Figure 2> Psychological Well-Being of Test Characteristic Curves for Korean and American Adults



<Figure 3> Psychological Well-Being of Test Characteristic Curves for Korean and American Adults

than in Korea and the levels at which most information occurs reflect higher trait levels for the former than the latter. For example, the test information curves (TIC) of Autonomy across the two nations are also presented in <Figure 2>, where the TIC for American adults is considerably steeper than that for Korean, indicating that the test provides more information for American adults. However, Purpose in Life shows different patterns with other dimensions. It has a small amount of information for both nations. In other words, Ryff's psychological well-being measurement

provides better information about the five dimensions for American adults than for Korean adults except Purpose in Life. However, the test characteristic curves (TCC) show an interesting result that the test provides more information for Korean rather than American adults.

## V. Conclusion

Does unidimensionality exist for the Psychological well-being scales for Korean and

American adults? And do factorial or invariance and item invariance exist for adults across the two nations? Findings from the present study indicate that the answer to these questions is a qualified no. That is, tests that examine invariance at factorial levels revealed disappointing results, where factorial variance was evident. This finding was also reflected in the results from IRT analyses which indicate that only seven items were invariant for all dimensions considered together and among these invariant items six provided enough psychometric information to meet our criterion for retention. In other words not only does the measure provide limited psychometric information for adults in both Korea and the United States, but also for the most part it measures the function of such adults on very different scales. Thus it is difficult to interpret findings from studies that survey and compare adults in two different nations, as the findings might reflect artifacts of measurement and not the true differences or similarities (Reise et al., 1993).

Measurement precision as indicated in the TICs and the amount of psychometric information evident in the measure across both nations is also noteworthy. Our analysis reveals that the different dimensions of the measure provide considerably more measurement precision for American adults than their Korean counterparts. These findings could support Diener et al.'s (1998) assumption that measures of subjective well-being such as the psychological well-being measure studied here might be weighted toward western values and in their present form such measures might not be culturally relevant for Koreans. Nevertheless it is difficult to judge whether the measure is more

suitable for American adults when it performs almost as poorly for American adults as it does for its Korean counterparts.

Shifting our focus to the level of functioning of each item on each dimension measures, it is important to note that <Table 3> indicates that across both nations, different items measure very different levels of functioning. When we simply sum up the rating to each item response adults from both nations, as it could lead to misleading results. Examining the pattern of responses a priori or a posteriori using IRT methodology might provide more accurate results.

It is important to interpret the present findings in the context of this study's limitations. One of the most common drawbacks researchers point out is the sample size. The sample used in this study is larger than that often recommended for IRT study (see Embretson and Reise, 2000) and those used in recent DIF studies (e.g., Cooke, et al., 2001; Orlando & Marshall, 2002). One important shortcoming is that regarding unidimensionality. That is, despite our examination of multiple factor solutions we could not find a factor model with factors that met established criteria for unidimensionality. While we used the existing model that more closely approximated the unidimensionality criteria, we note that its factors did not fully satisfy the unidimensionality criterion. Thus, we did not meet one of the most critical assumptions of measurement and especially that of IRT. Problems with unidimensionality can lead to inaccurate item parameter estimates. The persistence of this problem, however indicate that the measure might provide less than adequate psychometric indices for the constructs it purports

to measure.

Having discussed the studies limitations, we now turn to recommendations.

First, because of the limited amount of psychometric information most items on each dimension provides, it is difficult to endorse the continued use of the measure for assessment of adults of both nations in its present form. We therefore recommend the expenditure of sufficient effort to improve its psychometric properties. Because we cannot recommend its usage in either country and also because of the psychometric variance across both nations, we also cannot confirm its continued use across both nations in its present form. In conclusion we note that this study represents the first effort we know that has used the arguably more rigorous IRT methodology to address the psychometric properties of the Psychological Well-being measure cross-nationally. We hope that having provided cross-national parameter estimates for items on each dimension we have now provide a foundation for further research to improve on the psychometric properties of this existing measure.

## ■ References

- Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and presumption of equivalent measurement and theoretical structure: a look beneath the surface. *Journal of Cross-Cultural Psychology*, 30, 555-574.
- Clarke, P. J., Marshall, V. W., Ryff, C. D., & Wheaton, B. (2001). Measuring psychological well-being in the Canadian Study of Health and Aging. *International Psychogeriatrics*, 13, 79-90.
- Clarke, P. J., Marshall, V. W., Ryff, C. D., Rosenthal, C. J. (2000). Well being in Canadian seniors: Findings from the Canadian Study of Health and Aging. *Canadian Journal on Aging*, 19(2), 139-159
- Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist-Revised. *Psychological Assessment*, 9, 3-14.
- Cooke, D. J., & Michie, C. (1999). Psychopathy across cultures: North America and Scotland compared *Journal of Abnormal Psychology*, 108, 58-68.
- Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and Ethnicity: Structural item, and test generalization of the Psychopathy Checklist-Revised (PCLR) in Caucasian and African American participants. *Psychological Assessment*, 13, 531-542.
- De Ayala, R. J. (1993). An Introduction to Polytomous Item Response Theory Models. *Measurement and Evaluation in Counseling and Development*, 25, 172-89.
- DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods*, 3, 412-423.
- Diener, E., Sapyta, J. J., & Suh, E. M. (1998). Subjective well-being is essential to well-being. *Psychological Inquiry*, 28, 195-223.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gray-Little, B., Williams, V.S.L, & Hancock, T.D.

- (1998). An item response theory assessment of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 23, 443-451.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Keyes, C. L. M., & Ryff, C. D. (1998). Generativity in adult lives: Social structural contours and quality of life consequences. In D. P. McAdams & E. de St. Aubin (Eds.), *Generativity and adult development: How and why we care for the next generation* 227-264. Washington: American Psychology Association.
- Lambert, M.C., Schmitt, N., Samms-Vaughan, M.E., An, J.S., Fairclough, M., & Nutter C. (2003). Is it prudent to administer all items for each Child behavior Checklist Syndrome? Evaluating the psychometric properties of the Youth Self-Report dimensions with confirmatory factor analysis and item response theory. *Psychological Assessment*, 15, 550-568.
- Lu, L., Gilmour, R., & Kao, S. (2001). Cultural values and happiness: An East-West dialogue. *Journal of Social Psychology*, 141(4), 477-493.
- Mellenbergh, G. (1996). Measurement precision in test score and item response models. *Psychological Methods*, 1, 293-299.
- Orlando, M., & Marshall, G. N. (2002). Differential Item functioning in a Spanish translation of the PTSD checklist: Detection and evaluation of impact. *Psychological Assessment*, 14, 50-59.
- Reise, S.P. (1999). Personality measurement issues viewed through the eyes of IRT. In S.E. Embretson & S.L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and every Psychologist should know* 219-241. Mahwah, NJ: Lawrence Erlbaum Associates.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Romero, A. J. (2000). Assessing and treating Latinos: Overview of research. In I. Cuella & F. A. Paniagua. (Eds.), *Handbook of Multicultural Mental Health*. (pp.209-223). Sandiego: Academic Press.
- Ruini, C. O. F., Rafanelli, C., Ryff, C., & Fava, G. A. (2003). Italian validation of Psychological Well-being Scales (PWB). *Rivista di Psichiatria*, 38(3), 117-130.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well being. *Journal of Personality & Social Psychology*, 57(6), 1069-1081. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science*, 4(4), 99-104.
- Ryff, C. D. & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality & Social Psychology*, 69(4), 719-727.



- Ryff, C. D. & Heidrich, S. M. (1997). Experience and well-being: Explorations on domains of life and how they matter. *International Journal of Behavioral Development*, 20(2), 193-206
- Ryff, C., Lee, Y. H., & Na, K. C. (1993). Through the lens of culture: Psychological well-being at midlife. Paper presented at the meetings of Gerontological Society of America, New Orleans.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4, Pt. 2), 100.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum. Scientific Software International, Inc., (2002)
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Thissen, D., Steinberg, L., & Wainer, H. (1991). *MULTILOG users guide: Multiple categorical item analysis and test scoring using item response theory*. Chicago, IL: Scientific Software International.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research methods*, 3, 4-69.
- Weiss, D. J. (1995). Improving individual differences measurement with item response theory and computerized adaptive testing. In D. J. Lubinski, & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* 49-79. CA: Davies-Black Publishing.

---

Received October 10, 2004

Accepted November 29, 2004