### Title

A psychometric revision of the Asian values scale using the Rasch model




### Journal

Measurement and Evaluation in Counseling and Development, 37(1)

### ISSN

0748-1756



### Authors

Kim, Bryan S. K.
Hong, Sehee



### Publication Date

2004-04-01

Peer reviewed

# A Psychometric Revision of the Asian Values Scale Using the Rasch Model

Bryan S. K. Kim and Sehee Hong

*The 36-item Asian Values Scale (B. S. K. Kim, D. R. Atkinson, & P. H. Yang, 1999) was revised on the basis of G. Rasch's (1960) model and data from 618 Asian Americans. The results led to the establishment of a 25-item measure named the Asian Values Scale–Revised.*

Asian Americans represent one of the fastest growing and highly diverse ethnic groups in the United States (U.S. Bureau of the Census, 2002). Between 1980 and 1990, the number of Asian Americans in the United States doubled, mainly as a result of the Immigration Act of 1965 and the U.S. military involvement in Southeast Asia. As of 2000, there were nearly 12 million Asian Americans, representing an increase of 72% since 1990 and a figure of 4.2% of the current U.S. population (U.S. Bureau of the Census, 2002). It has been projected that by 2050 approximately 1 out of 10 Americans will have an Asian ancestry (U.S. Bureau of the Census, 1992). In terms of diversity within the Asian American group, the U.S. Bureau of the Census (2002) has identified no less than 24 different ethnic groups, including Asian Indian, Chinese, Japanese, Filipino, Korean, Laotian, and Vietnamese. Each of these groups has its own unique language, immigration history, traditions, and customs.

As a result of these growing numbers and the recognition of the high degree of within-group variability, scholars have focused on identifying psychological constructs that are salient to the experiences of Asian Americans. These efforts have led to the concept of *enculturation*. First defined and used by Herskovits (1948), enculturation refers to the process of socialization to the norms of one's indigenous culture, including the values, ideas, and concepts that are salient for the culture (Berry, Poortinga, Segall, & Dasen, 1992; Segall, Dasen, Berry, & Poortinga, 1990). Recently, enculturation was defined as the process of retaining one's indigenous cultural values, behaviors, knowledge, and identity (Kim & Abreu, 2001; Kim, Atkinson, & Umemoto, 2001). In particular, Kim, Atkinson, et al. (2001) noted that an important dimension of enculturation for Asian Americans is adherence to Asian cultural values; cultural values refer to "universalistic statements about what we think are desirable or attractive" (P. B. Smith & Bond, 1994, p. 52). Asian cultural values that are salient for Asian Americans include collectivism, conformity to norms, deference to authority figures, emotional restraint, filial piety, hierarchical family structure, and humility (Kim, Atkinson, & Yang, 1999; Sue & Sue, 1999). Current theories on Asian cultural values suggest that first-generation Asian Americans adhere to these values more strongly than Asian Americans who are many generations removed from immigration (Kim, Atkinson, et al., 2001). In addition, the theories suggest that adherence to these values influences the ways in which individuals manifest psychological problems, people's beliefs about problem etiology, people's help-seeking behaviors, and the ways in which individuals express their emotions (Atkinson, Morten, & Sue, 1998; Sue & Sue, 1999).

*Bryan S. K. Kim, Counseling, Clinical, and School Psychology Program, Department of Education, University of California, Santa Barbara; Sehee Hong, Department of Psychology, Ewha Womans University, Seoul, Republic of Korea. Correspondence concerning this article should be addressed to Bryan S. K. Kim, Counseling, Clinical, and School Psychology Program, Department of Education, University of California, Santa Barbara, CA 93106-9490 (e-mail: bkim@education.ucsb.edu).*

To empirically examine the construct of Asian cultural values and its relationships to other psychological concepts, Kim et al. (1999) developed the Asian Values Scale (AVS). The AVS contains 36 items and uses a 7-point Likert-type scale (1 = *strongly disagree,* 2 = *moderately disagree,* 3 = *mildly disagree,* 4 = *neither agree nor disagree,* 5 = *mildly agree,* 6 = *moderately agree,* and 7 = *strongly agree*). Example items are (a) "One should be discouraged from talking about one's accomplishments," (b) "The worst thing one can do is to bring disgrace to one's family reputation," (c) "One should consider the needs of others before considering one's own needs," (d) "One should be humble and modest," and (e) "One need not be able to resolve psychological problems on one's own" (reverse worded). For the 36-item scale's score, Kim et al. (1999) reported coefficient alphas of .81 and .82 and a 2-week test–retest reliability coefficient of .83. Support for the AVS score's construct validity was obtained by identifying, via a nationwide survey and focus-group discussions, items that reflect cultural values commonly observed across various Asian American ethnic groups; items were retained that were more highly endorsed by first-generation Asian Americans than by European Americans. Evidence of the AVS score's concurrent validity was obtained through a confirmatory factor analysis, in which a factor structure comprising the AVS, the Individualism–Collectivism Scale (Triandis, 1995), and the Suinn–Lew Asian Self-Identity Acculturation Scale (SL-ASIA; Suinn, Rickard-Figueroa, Lew, & Vigil, 1987) was confirmed. Discriminant validity was evidenced in the low correlation between the AVS scores, which reflect values enculturation, and the SL-ASIA scores, which reflect predominantly behavioral acculturation.

The AVS has been used in a number of research studies, mostly in the field of counseling process and outcome. In particular, two studies have examined the effects of client adherence to Asian cultural values on single-session outcome with Asian American college student clients. Kim and Atkinson (2002) found that when counselors were Asian American, the clients who had high scores on the AVS rated the counselors as more empathic and credible than did the clients with low AVS scores. In contrast, when the counselors were European American, the clients with low AVS scores rated the counselors as more empathic than did the clients with high AVS scores. However, in another study of single-session counseling, Kim, Li, and Liang (2002), using all European American counselors, found that Asian American clients with high scores on the AVS perceived greater counselor empathy and working alliances than did clients with low AVS scores. This latter finding is directly contradictory to the findings from Kim and Atkinson's (2002) sample with European American counselors. It is difficult to make sense of the inconsistent findings regarding values adherence and differences in counselor ethnicity, because there may be many alternative explanations for these findings (e.g., presence of moderator variables). However, one possibility is that the inconsistent results might reflect some psychometric limitation with the AVS in its assessment of adherence to cultural values.

Although the AVS score has evidence of reliability and validity on the basis of classical test theory, each of the 36 items has not been subjected to a more rigorous examination. Because the items were chosen simply on the basis of *t* tests of group-level scores between first-generation Asian Americans and general European Americans, the functional characteristics of each item are not known. It would be useful to have information on the extent to which each item is difficult to endorse; some AVS items may be more difficult to endorse than are other items. For example, in relation to the value of humility, one could reason that the AVS item "One should be discouraged from talking about one's accomplishments" could be more difficult to endorse than the item "One should be humble and modest" because the former item is more behaviorally specific than the latter one. In addition, it would be helpful to know whether the difficulty levels of the items reflect the full range of respondents' trait levels. The AVS items may be such that respondents with a high level of adherence to Asian values may not be fully distinguishable from respondents with only a moderate level of adherence to the values.

In addition to item characteristics, it would be useful to know how well the 7-point scale expresses the distinctions between each category of agreement. To be more specific, it is unclear how well the 7-point scale is able to differentiate between individuals who score a 2 *(moderately disagree)* and those who score a 3 *(mildly disagree)* or between persons who score a 5 *(mildly agree)* and those who score a 6 *(moderately agree).* In addition to distinctions between levels of agreement, it is unclear whether the score of 4 *(neither agree nor disagree)* accurately reflects the halfway point between *strongly disagree* and *strongly agree* or represents an altogether different construct such as item irrelevance.

Hence, our goal in the present study was to conduct a more rigorous psychometric analysis of the AVS items. For this analysis, we used Rasch's (1960) model. In doing so, we sought to (a) study whether the 7-point categories are appropriate for the AVS items, (b) remove any misfitting items by investigating the extent to which each of the 36 AVS items supports unidimensionality, and (c) investigate the appropriateness of the remaining items by assessing the relative distributions of the person-trait-level and item-difficulty-level estimates on a common logit scale.

## METHOD

### Participants

Respondents were 618 (303 male, 315 female) Asian American college students at universities in California, Hawaii, and Maryland. They ranged in age from 18 to 37 years, with a mean of 21.0 years (*SD* = 4.7). Two hundred seven (33.5%) were freshmen, 145 (23.5%) were seniors, 110 (17.8%) were sophomores, 99 (16.0%) were juniors, and 55 (8.9%) were graduate students; 2 (0.3%) did not report grade level. In terms of ethnicity, there were 152 (24.6%) Chinese, 136 (22.0%) Koreans, 86 (13.9%) Filipinos, 64 (10.4%) Japanese, 51 (8.3%) Asian Indians, 29 (4.7%) multiethnic Asians, 26 (4.2%) Vietnamese, 15 (2.4%) Taiwanese, 14 (2.3%) multiracial Asians, and 42 (6.8%) other Asian Americans; 3 (0.5%) did not report ethnicity. In terms of generations since immigration, 277 (44.8%) were second generation, 239 (38.7%) were first generation, 40 (6.5%) were third generation, 33 (5.3%) were fourth generation, 17 (2.8%) were fifth generation, and 3 (0.5%) were sixth generation; 9 (1.5%) did not report generation status.

### Procedure

Data were obtained from past studies that incorporated the AVS as one of the measures of the study. These studies were Chung, Kim, and Abreu (2004); Kim and Atkinson (2002); Kim et al. (1999); Kim et al. (2002); Kim, Hill, et al. (2003); Kim, Li, and Ng (2003); Kim and Omizo (2003); and Liang, Li, and Kim (2004).

### Data Analysis

The basic Rasch model is a dichotomous response model (Wright & Stone, 1979), which represents the conditional probability of a binary outcome as functions of a person's trait level *(B)* and an item's difficulty level *(D):*

$$P(x = 1) = \frac{\exp(B_n - D_i)}{1 + \exp(B_n - D_i)},$$

where $P(x = 1)$ is the probability of an endorsed response ("yes" response to an item), $B_n$ is the trait parameter of person *n,* and $D_i$ is the difficulty of endorsing item *i*. When $B_n > D_i$, $B_n = D_i$, and $B_n < D_i$, the chance of a "yes" response is greater than 50%, equal to 50%, and less

than 50%, respectively. Given that trait levels and item difficulty levels are placed on a common metric (logit), a major advantage of the Rasch model is the direct comparison of these two parameters. In the Rasch model, all items are assumed to have equal discriminating power. Nevertheless, the Rasch model has been found to be robust to departures from the equal discrimination assumption (Dinero & Haertel, 1977). One of the most important theoretical merits of the Rasch model is its "specific objectivity" (Rasch, 1977), which means that the comparison of the difficulty of two items should not depend on the ability levels of the persons used to measure the difficulty on the items and, also, that the comparison of the ability of two persons should not depend on the difficulties of the items used to measure the ability of the persons.

The Rasch model can be generalized to polytomous items with ordered categories. The formulation of an extended Rasch model includes the partial credit model (PCM; Masters, 1982) and the rating scale model (RSM; Andrich, 1978). The PCM was originally developed for analyzing achievement test items that include multiple solving steps, in which partial credit can be assigned for completing several steps in the solution process (Embretson & Reise, 2000). In addition, the PCM is also useful for analyzing attitude or personality scale responses. Assuming that item $i$ is scored $x = 1, 2, \ldots, m$ for $x = j$, the PCM can be expressed mathematically as

$$P_{xni} = \frac{\exp[\sum_{j=1}^{x} (B_n - \delta_{ij})]}{\sum_{k=1}^{m_i} [\exp \sum_{j=1}^{k} (B_n - \delta_{ij})]},$$

where the $\delta_{ij}$ term is called the item step difficulty associated with category $j$. In the PCM, the $\delta_{ij}$ parameter represents the relative difficulty of each step. Thus, the higher the value of a particular $\delta_{ij}$, the greater the difficulty level of a particular step relative to other steps within an item (Embretson & Reise, 2000). In contrast, the RSM is a subset of the PCM because it restricts the step structure to being the same for all items (Wright & Masters, 1982). Thus, in the RSM, a common set of $\delta_{ij}$ parameters is estimated. The parameter $\delta_{ij}$, step calibration, is also known as the threshold (Andrich, 1978). Thus, the RSM is useful when psychological distances between categories are the same for all items. Given that Likert scales can be modeled according to either a PCM or an RSM, it is necessary to determine which polytomous Rasch model and its respective set of estimated parameters best explain the data. To choose an appropriate model, several estimates obtained from the PCM and the RSM should be compared.

However, before we proceeded with the Rasch analysis, we first needed to evaluate the dimensionality of the AVS items because unidimensionality is considered the most critical and basic assumption of Rasch models. Hence, we performed an exploratory factor analysis for ordered polytomous data by using the unweighted least squares method available in Mplus (Muthen & Muthen, 1998). The items can be said to be roughly unidimensional if the first eigenvalue is relatively large in comparison to the second eigenvalue and if the second eigenvalue is not much larger than any of the other eigenvalues (Hambleton & Traub, 1973; Lord, 1980). The results yielded the following eigenvalues for the first, second, and third factors, respectively: 7.0, 2.9, and 1.9. These results support the Rasch analysis requirement for unidimensionality. It is important to note that although Kim et al. (1999) reported a six-factor structure for the AVS, a subsequent study by Kim, Yang, Atkinson, Wolfe, and Hong (2001) found that AVS items were best represented by a hierarchical structure with a unidimensional general Asian values factor on the second order followed by six factors on the first order.

Having satisfied the unidimensionality assumption, we compared the two polytomous Rasch models by first investigating whether scoring category transitions remained similar across all of the items. If the probability of stepping from one category to the next was roughly the same across all items, then the RSM would be the model of choice (Ludlow & Haley, 1992). The results with the PCM indicated that the scoring category transitions across many items were similar enough to support the selection of the RSM. In addition, the RSM results indicated that the scoring category transitions remained the same across all items. It should be noted that the Rasch analysis was undertaken using the WINSTEPS program (Linacre & Wright, 1999).

Next, we compared the Rasch reliability indices obtained from the two models. The Rasch analogue to Cronbach's alpha is called "person separation reliability" (Wright & Masters, 1982), which refers to the ability to differentiate persons on the measured variable and the replicability of person placement across other items measuring the same construct. The index ranges from 0 to 1, with values equal to or greater than .80 being regarded as acceptable (Fox & Jones, 1998). The person separation reliability is .86 for the PCM and .86 for the RSM, both of which are acceptable.

We also compared the scores of item separation reliability, which refers to the ability to define a distinct hierarchy of items along the measured variable and the replicability of item placement within the hierarchy across other samples. The item separation reliability is .99 for the PCM and .99 for the RSM, indicating that the reliability indices are again acceptable for both models.

On the basis of these results, we used the RSM for the present study. To summarize the rationale for this choice, first, the results show that all of the AVS items demonstrate the same rating scale with no evidence that the rating scale is differentially used across items. Second, given that the two models produced comparable person and item reliability estimates, a more parsimonious model (i.e., RSM) is preferred. Over-parameterizing should be avoided to account for idiosyncrasies of the data (Bode, 2001). Note that the PCM increases the number of additional free parameters to be estimated by $(L - 1)(m - 2)$, where $L$ is the number of items and $m$ is the number of categories in the rating scale.

## RESULTS

### Test of Category Use

It has been suggested that evaluating how respondents use the rating scale is the first step in conducting rating scale analysis (Lopez, 1996). It is often the case that respondents fail to react to a rating scale in the manner in which the test constructor intended (Roberts, 1994). Because it is always uncertain how a rating scale was used by individuals in a sample, an investigation of the functioning of the rating scale is always necessary (Linacre, 2002) and can be done with the Rasch analysis. In the Rasch analysis, a useful diagnostic in evaluating category usage is to examine the average measure and threshold of each category. The average measures across categories represent the empirical averages of the measure (Rasch trait or ability score) that are modeled to produce the responses observed in each category (Linacre & Wright, 1998). Because observations in higher categories must be produced by higher measures, the average measures across categories must increase monotonically. The thresholds indicate the measures at which adjacent categories are equally probable and thus define the boundaries between the categories. Therefore, the thresholds too should increase monotonically.

For the 7-point scale of the 36-item AVS, the average measure increased with the category label (–0.39, –0.25, –0.13, 0.04, 0.19, 0.32, and 0.52 for Categories 1 to 7, respectively), suggesting that the rating scale categorization is satisfactory. However, contrary to expectations, threshold estimates were not ordered, with logits of –0.68, –0.41, 0.07, –0.20, 0.34, and 0.87 for Categories 2 to 7, respectively; the lack of ordered increase occurs at Category 4.

Failure of threshold parameters to increase monotonically is known as "step disordering" (Linacre, 2002). Step disordering indicates a low probability of observance of certain categories. In addition, an examination of the probability curves for the seven categories showed that Category 4 is never modal, indicating that at no point on the continuum is Category 4 ever the most likely category to be observed. That is, regardless of a person's trait level, the probability of choosing Category 4 is never the most likely.

On the basis of the threshold results, we conducted a follow-up Rasch analysis to identify the optimal categorization. First, the neutral category was collapsed with the agree side (Category 5) and then with the disagree side (Category 3) of the scale to see if the threshold increased monotonically for either of the changes. Specifically, the rescoring methods were 1234456 and 1233456, respectively. For the rescoring method of 1234456, the original categories of 1, 2, and 3 were retained, but the original categories of 4 and 5 were collapsed into Category 4. Then, the original categories of 6 and 7 were rescored as 5 and 6, respectively. A similar procedure was applied to the rescoring method of 1233456. However, both rescoring methods failed to achieve step ordering. In addition, other scoring schemas (e.g., 1233345) similarly failed to produce monotonically increasing thresholds. Hence, we concluded that the neutral category might be a poorly defined category in the minds of the participants and decided to omit the category by treating it as missing data. This decision to eliminate the neutral category was supported by several studies that showed a neutral category often does not work as intended (Andrich, 1982; Bock & Jones, 1968; Dubois & Burns, 1975).

The rating scale without a neutral category, 123×456, where × indicates the category that was treated as missing, resulted in monotonically increasing thresholds (with logits of –0.78, –0.42, –0.16, 0.36, and 0.99 for Categories 2 to 6, respectively). However, this revised scoring still was not satisfactory. When the probability curves for the six categories (i.e., without Category 4) were examined, the segment on the continuum represented by each of the six categories was unacceptably narrow.

Consequently, Categories 2 and 3 and Categories 4 and 5 were combined, leading to the revised scoring method of 122×334. When this scoring method was used, the four category thresholds increased monotonically at approximately equal increments (with logits of –1.82, –0.19, and 2.01 for Categories 2 to 4), suggesting support for this scoring method; the average measures were –0.90, –0.40, 0.51, and 1.14 for Categories 1 to 4, respectively. In further support of this scoring method, the corresponding probability curves as shown in Figure 1 exhibited the desired appearance of a range of hills. Hence, on the basis of these results, we determined that the four-category scoring method (1 = *strongly disagree,* 2 = *disagree,* 3 = *agree,* and 4 = *strongly agree*) is appropriate for the AVS items.

## Dimensionality

To study whether the data supported a unidimensional structure, we computed item fit mean square (MNSQ) values by using the RSM. MNSQ determines how well each item contributes to defining one common construct. Item MNSQ values of about 1.0 are ideal by Rasch specifications, and values greater than 1.2 and less than 0.8 for samples greater than 500 are unacceptable (R. M. Smith, Schumacker, & Bush, 1995). The cutoff values, however, tend to vary depending on the purpose for which the ratings are used (Karabatsos, 1997). High values of item MNSQ indicate a lack of construct homogeneity with other items in a scale, whereas low values indicate redundancy with other items (Linacre & Wright, 1998). Two MNSQ statistics were used: *infit* (weighted) and *outfit* (unweighted) statistics. It is important to note that although infit and outfit statistics do not provide a complete dimensionality test, they nonetheless are related to, and provide important information about, dimensionality.

On the basis of the items' MNSQ values, most of the items indicated reasonable fit to the RSM. However, AVS Items 1, 6, 18, 22, 25, 34, and 36 revealed both infit and outfit statistics greater than 1.2. AVS Items 9, 20, 29, and 30 revealed both infit and outfit statistics less than

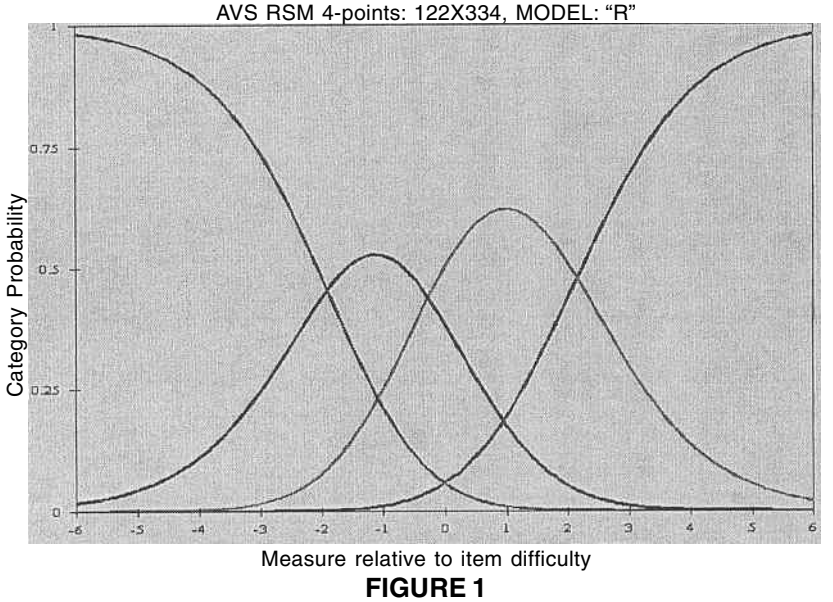


**FIGURE 1**

**Response Functions for Four Categories**

0.8. Hence, these 11 misfitting items were deleted. Another RSM analysis with the remaining 25 items was performed to test the item fit. As shown in Table 1, no items, except for Item 3, produced infit and outfit values greater than 1.2 and less than 0.8. The infit and outfit values of Item 3 were very close to 1.2, and thus all of the items showed reasonable fit to the RSM.

Given this change in the number of AVS items, we conducted additional analyses on the appropriateness of the previously determined 4-point scale for the remaining 25 items. The results showed that the thresholds of the four categories were monotonically increasing (–1.94, –0.19, and 2.13 for Categories 2 to 4) and that there were nearly equal distances between categories (see Figure 2); the average measures were –1.07, –0.50, 0.46, and 1.00 for Categories 1 to 4. In addition to these analyses, the 25-item scale showed a person separation reliability of .80. As we mentioned earlier, this index ranges from 0 to 1, with values equal to or greater than .80 regarded as acceptable (Fox & Jones, 1998).

## Appropriateness of the AVS Difficulty Level for Samples

As we stated earlier, the Rasch model estimates person trait and item difficulty parameters on a common metric, that is, logits. Thus, it is possible to test whether the difficulty level of the AVS is appropriate for the sample in this study. If an instrument is appropriately targeted for the sample being tested, there should be much overlap between the range of the person trait levels and that of item difficulty levels. A distribution of person trait level and item difficulty on a logit scale for the 25 items and using the 4-point scale is shown in Figure 3. The left side of the continuum illustrates the person trait level, and the right side of the scale shows the item difficulty level. As shown, there is considerable overlap on both sides of the scale. The coverage of the 25 items, which spans the entire range of a person's ability, indicates that this set of items is appropriately targeted for the sample being measured.

In summary, the AVS was revised to now include 25 items and a 4-point Likert-type scale (1 = *strongly disagree,* 2 = *disagree,* 3 = *agree,* and 4 = *strongly agree*). This revised instrument was named the Asian Values Scale–Revised (AVS-R) and is available from Bryan S. K. Kim.

## TABLE 1

### Parameter Values for the Remaining 25 Items of the Asian Values Scale (in Order of Item Difficulty)

| Item Number and Item | Infit MNSQ | Outfit MNSQ | Difficulty |
|---|---|---|---|
| 15 One should be able to question a person in an authority position.* | 0.87 | 0.86 | 1.39 |
| 11 One need not minimize or depreciate one's own achievements.* | 0.89 | 0.89 | 1.14 |
| 7 Younger persons should be able to confront their elders.* | 0.97 | 0.97 | 1.03 |
| 24 One need not remain reserved and tranquil.* | 0.96 | 0.96 | 0.80 |
| 4 One need not focus all energies on one's studies.* | 1.10 | 1.12 | 0.75 |
| 28 One need not be able to resolve psychological problems on one's own.* | 1.14 | 1.15 | 0.63 |
| 33 One should not make waves. | 1.07 | 1.07 | 0.62 |
| 5 One should be discouraged from talking about one's accomplishments. | 1.05 | 1.07 | 0.55 |
| 32 One need not follow the role expectations (gender, family hierarchy) of one's family.* | 0.85 | 0.85 | 0.50 |
| 10 One need not achieve academically in order to make one's parents proud.* | 1.09 | 1.10 | 0.29 |
| 27 Family's reputation is not the primary social concern.* | 0.85 | 0.86 | 0.25 |
| 2 One should not deviate from familial and social norms. | 0.84 | 0.84 | 0.13 |
| 23 The worst thing one can do is to bring disgrace to one's family reputation. | 1.11 | 1.11 | 0.11 |
| 14 One should think about one's group before oneself. | 0.97 | 0.97 | –0.09 |
| 31 Occupational failure does not bring shame to the family.* | 0.98 | 0.97 | –0.17 |
| 17 One's achievements should be viewed as family's achievements. | 0.95 | 0.94 | –0.17 |
| 13 Educational and career achievements need not be one's top priority.* | 1.19 | 1.19 | –0.18 |
| 35 One need not control one's expression of emotions.* | 1.12 | 1.12 | –0.49 |
| 8 When one receives a gift, one should reciprocate with a gift of equal or greater value. | 1.08 | 1.07 | –0.54 |
| 12 One should consider the needs of others before considering one's own needs. | 0.91 | 0.90 | –0.60 |
| 21 One should have sufficient inner resources to resolve emotional problems. | 0.89 | 0.87 | –0.81 |
| 19 One should avoid bringing displeasure to one's ancestors. | 0.93 | 0.91 | –0.86 |
| 3 Children should not place their parents in retirement homes. | 1.22 | 1.22 | –1.30 |
| 26 One should be humble and modest. | 0.93 | 0.93 | –1.37 |
| 16 Modesty is an important quality for a person. | 0.90 | 0.91 | –1.60 |

*Note*. MNSQ = mean square. Asterisk indicates reverse-worded item.

## DISCUSSION

The AVS was designed to assess a person's adherence to Asian cultural values. On the basis of classical test theory, Kim et al. (1999) reported evidence of the reliability and validity of AVS scores. In an attempt to improve the measurement quality of the scale, in the present study we used the Rasch model to revise the AVS. As a result, the 25-item AVS-R, with a 4-point scale, was established.

The first step in establishing the AVS-R was to investigate the adequacy of the 7-point scale of the original AVS. The results supported the deletion of three categories and led to a 4-point scale. The results suggested that Category 4 might be poorly defined by respondents or might reflect a construct that is unrelated to the disagree–agree continuum. In addition to Category 4, Categories 3 *(mildly disagree)* and 5 *(mildly agree)* were also found to be inadequate in fully representing participant responses and hence were deleted. Although these categories might have been used originally because they were assumed to be

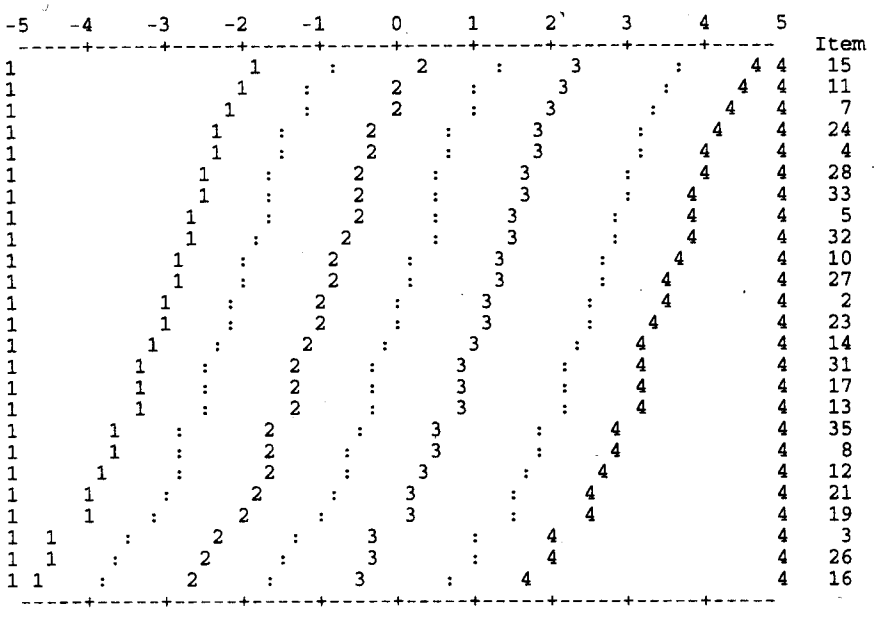


**FIGURE 2**

**Estimated Scoring Category Transitions Using RSM**

*Note.* RSM = rating scale model.

able to lead to a more precise index of agreement, the present findings suggest that these categories may instead lead to definitional confusion on the part of the respondents. Furthermore, the lack of support for the assumption of equal spacing between the categories (e.g., between *mildly agree* and *moderately agree* and between *moderately agree* and *strongly agree*) suggests that the use of adjectives such as *mildly* and *moderately* does not accurately represent equal levels of intensity (i.e., increase of thresholds at equal increments).

The second step in establishing the AVS-R involved identifying misfitting items on the basis of infit and outfit statistics. Among the 11 items that were removed from the scale, 7 indicated lack of construct homogeneity, and 4 represented redundant items. Some of the items that indicated lack of construct homogeneity were "Children need not take care of their parents when the parents become unable to take care of themselves" (reverse worded), "One's family need not be the main source of trust and dependence" (reverse worded), and "Parental love should be implicitly understood and not openly expressed." Although these items were written on the basis of the existing literature on Asian cultural values, the results suggest that the items represent different constructs in comparison to the construct assessed by the AVS-R's 25 items. Among the redundant items, 3 were as follows: "Following familial and social expectations are [*sic*] important," "One need not conform to one's family's and the society's expectations" (reverse worded), and "One need not follow one's family's and the society's norms" (reverse worded). All of these items are variations of the following item that was retained in the AVS-R: "One should not deviate from familial and social norms." Hence, this comparison supports the conclusion that the deleted items were redundant. It is interesting to note that despite the elimination of about 30% of the original AVS items, the AVS-R score still retained adequate reliability (i.e., person separation reliability of .80); Kim et al. (1999) reported internal consistency coefficients of .81 and .82 for the original 36-item AVS. This finding further supports the conclusion that the deleted items do not contribute

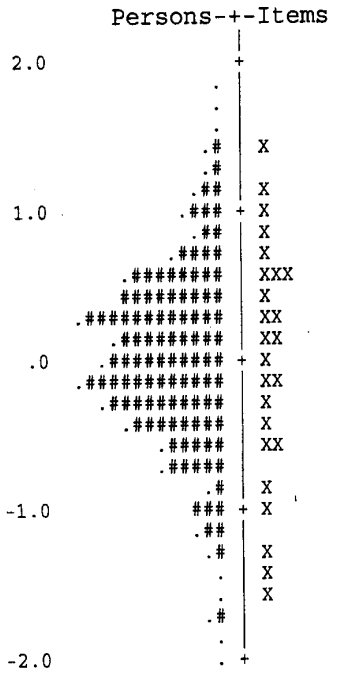


**FIGURE 3**

**A Distribution of Person Trait Level and Item Difficulty Level for 25 Items**

*Note*. Each "#" represents two participants and each "." represents one person. Each "×" indicates one item.

to assessing a homogeneous construct and do not add related information to the scale. In sum, the AVS-R represents a streamlined version of the AVS, without sacrificing reliability.

The third and final step in establishing the AVS-R was to study the adequacy of the AVS-R through the use of a person–item map. The map indicated that in terms of difficulty level, the AVS-R items functioned well to represent the full range of respondents' trait abilities. Despite the elimination of 11 items, there was an adequate overlap between person trait and item difficulty levels, which suggests that the AVS-R is adequate to assess individuals at various levels of values adherence. As we mentioned earlier, it is important for an instrument to be able to assess both individuals with a high level of cultural values adherence and individuals with a low level of cultural values adherence. The lack of sensitivity in assessing individuals at either end of the trait-level continuum could lead to a ceiling or a flooring effect and the resulting inability of the instrument to detect the full variability in a population. In sum, the results of the person–item map provided additional evidence of the AVS-R's psychometric improvement.

As with any study, the present findings have limitations. First, the generalizability of the findings is limited to the Asian American college student population. Because the AVS-R was developed through the use of data from this population, the adequacy of the AVS-R with other Asian American populations should be examined before it is used with them. Second, although the data were obtained from the West Coast, the East Coast, and the Pacific, the results may not generalize to Asian Americans living in other areas. Third, although the present sample of Asian Americans included many ethnic groups, the results are more salient for those groups overrepresented in the study and less salient for those underrepresented or not represented at all.

The results of the present study have implications for future research. First, this study should be replicated with another sample of Asian Americans. Although we used a relatively large sample in this study, it would be useful if the psychometric properties of the AVS-R were further studied in another sample to ensure that the Rasch model results in this study can be replicated. A future study should attempt to gather data from a large and ethnically and geographically diverse group of Asian Americans. Second, researchers might consider further studying the AVS-R by applying other polytomous item response theory models. For example, Samejima's (1972) grade response model, which takes into consideration not only the item difficulty but also the item discrimination, might be a useful model. Third, the AVS-R should be administered with other similar instruments to study the construct validity of AVS-R scores within the paradigm of classical test theory. Although the AVS-R is based on the AVS, the number of modifications that were made to establish the AVS-R warrants additional study of the validity of AVS-R scores. For the present data, we observed a Pearson correlation coefficient of .93 ($p = .000$) between the AVS and the AVS-R scores, suggesting concurrent validity for the AVS-R scores. In addition, in a future study researchers might examine possible differences between Asian American men and women on the AVS-R score as a way to study the criterion-related validity of the score.

In conclusion, the 25-item AVS-R represents a psychometric advancement of the original 36-item AVS. The successful application of the Rasch model led to a more streamlined instrument with increased quality. We hope that the availability of the AVS-R represents an improved instrumentation in the assessment of Asian Americans' adherence to Asian cultural values. We also hope that the availability of this instrument will promote increased research in the area of Asian cultural values among Asian Americans.

## REFERENCES


Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika, 43,* 561–573.

Andrich, D. (1982). Using latent trait measurement models to analyze attitudinal data: A synthesis of viewpoints. In D. Spearritt (Ed.), *Proceedings of the 1980 Invitational Conference for the Improvement of Testing* (pp. 89–126). Melbourne, Victoria, Australia: Australian Council for Educational Research.

Atkinson, D. R., Morten, G., & Sue, D. W. (1998). *Counseling American minorities* (5th ed.). Boston: McGraw-Hill.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology: Research and application.* New York: Cambridge University Press.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgement and choice.* San Francisco: Holden Day.

Bode, R. K. (2001). Partial credit model and pivot anchoring. *Journal of Applied Measurement, 2,* 78–95.

Chung, R., Kim, B. S. K., & Abreu, J. M. (2004). Asian American Multidimensional Acculturation Scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology, 10,* 66–80.

Dinero, T. E., & Haertel, E. (1977). The applicability of the Rasch model with varying item discriminations. *Applied Psychological Measurement, 1,* 581–593.

Dubois, B., & Burns, J. A. (1975). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement, 35,* 869–884.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45,* 30–45.

Hambleton, R. K., & Traub, R. (1973). Analysis of empirical data using two latent trait models. *British Journal of Mathematical and Statistical Psychology, 26,* 195–211.

Herskovits, M. J. (1948). *Man and his works: The science of cultural anthropology.* New York: Knopf.

Karabatsos, G. (1997). The Sexual Experiences Survey: Interpretation and validity. *Journal of Outcome Measurement, 1,* 305–328.

Kim, B. S. K., & Abreu, J. M. (2001). Acculturation measurement: Theory, current instruments, and future directions. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (2nd ed., pp. 394–424). Thousand Oaks, CA: Sage.

Kim, B. S. K., & Atkinson, D. R. (2002). Asian American client adherence to Asian cultural values, counselor expression of cultural values, counselor ethnicity, and career counseling process. *Journal of Counseling Psychology, 49,* 3–13.
Kim, B. S. K., Atkinson, D. R., & Umemoto, D. (2001). Asian cultural values and the counseling process: Current knowledge and directions for future research. *The Counseling Psychologist, 29,* 570–603.
Kim, B. S. K., Atkinson, D. R., & Yang, P. H. (1999). The Asian Values Scale: Development, factor analysis, validation, and reliability. *Journal of Counseling Psychology, 46,* 342–352.
Kim, B. S. K., Hill, C. E., Gelso, C. J., Goates, M. K., Asay, P. A., & Harbin, J. M. (2003). Counselor self-disclosure, East Asian American client adherence to Asian cultural values, and counseling process. *Journal of Counseling Psychology, 50,* 324–332.
Kim, B. S. K., Li, L. C., & Liang, C. T. H. (2002). Effects of Asian American client adherence to Asian cultural values, session goal, and counselor emphasis of client expression on career counseling process. *Journal of Counseling Psychology, 49,* 342–354.
Kim, B. S. K., Li, L. C., & Ng, G. F. (2003). *Asian Values Scale–Multidimensional: Development, reliability, and validity.* Manuscript submitted for publication.
Kim, B. S. K., & Omizo, M. M. (2003). Asian cultural values, attitudes toward seeking professional psychological help, and willingness to see a counselor. *The Counseling Psychologist, 31,* 343–361.
Kim, B. S. K., Yang, P. H., Atkinson, D. R., Wolfe, M. M., & Hong, S. (2001). Cultural value similarities and differences among Asian American ethnic groups. *Cultural Diversity and Ethnic Minority Psychology, 7,* 343–361.
Liang, C. T. H., Li, L. C., & Kim, B. S. K. (2004). The Asian American Racism-Related Stress Inventory: Development, factor analysis, reliability, and validity. *Journal of Counseling Psychology, 51,* 103–114.
Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3,* 85–106.
Linacre, J. M., & Wright, B. D. (1998). *A user's guide to WINSTEPS.* Chicago: MESA Press.
Linacre, J. M., & Wright, B. D. (1999). WINSTEPS: Rasch analysis for all two-facet models (Version 2.98) [Computer software]. Chicago: MESA Press.
Lopez, W. (1996). Communication validity and rating scaler. *Rasch Measurement Transactions, 10,* 482–483.
Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.
Ludlow, L. H., & Haley, S. M. (1992). Polytomous Rasch models for behavioral assessment: The Tufts assessment of motor performance. In M. Wilson (Ed.), *Objective measurement I: Theory into practice* (pp. 121–137). Norwood, NJ: Ablex.
Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.
Muthen, L. K., & Muthen, B. O. (1998). *Mplus user's guide.* Los Angeles: Author.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: MESA Press.
Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In M. Glegvad (Ed.), *The Danish yearbook of philosophy* (pp. 58–94). Copenhagen, Denmark: Munksgaard.
Roberts, J. (1994). Rating scale functioning. *Rasch Measurement Transactions, 8,* 386.
Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph Supplement, 37*(1, Pt. 2).
Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1990). *Human behavior in global perspective: An introduction to cross-cultural psychology.* New York: Pergamon Press.
Smith, P. B., & Bond, M. H. (1994). *Social psychology across cultures: Analysis and perspectives.* London: Allyn & Bacon.
Smith, R. M., Schumacker, R. E., & Bush, J. M. (1995, April). *Using item mean squares to evaluate fit to the Rasch model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
Sue, D. W., & Sue, D. (1999). *Counseling the culturally diverse: Theory and practice* (3rd ed.). New York: Wiley.
Suinn, R. M., Rickard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn–Lew Asian Self-Identity Acculturation Scale: An initial report. *Educational and Psychological Measurement, 47,* 401–407.
Triandis, H. C. (1995). *Individualism and collectivism.* Boulder, CO: Westview Press.

U.S. Bureau of the Census. (1992). *Population projections of the United States, by age, sex, race, and Hispanic origin: 1992 to 2050* (Current Population Reports, P25-1092). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (2002). *Census brief: Current population survey.* Retrieved April 3, 2002, from http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.