

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

A psychometric revision of the European American Values Scale for Asian Americans using the Rasch model

Permalink

<https://escholarship.org/uc/item/3069m111>

Journal

Measurement and Evaluation in Counseling and Development, 37(4)

ISSN

0748-1756

Authors

Hong, S
Kim, Bryan S.K.
Wolfe, M M

Publication Date

2005

Peer reviewed

ARTICLES

A Psychometric Revision of the European American Values Scale for Asian Americans Using the Rasch Model

Sehee Hong, Bryan S. K. Kim, and Maren M. Wolfe

The 18-item European American Values Scale for Asian Americans (M. M. Wolfe, P. H. Yang, E. C. Wong, & D. R. Atkinson, 2001) was revised on the basis of results from a psychometric analysis using the Rasch Model (G. Rasch, 1960). The results led to the establishment of the 25-item European American Values Scale for Asian Americans—Revised.

Scholars have identified several salient psychological constructs related to the experiences of Asian Americans, an ethnic group that is one of the most diverse and fastest growing in the United States (Atkinson, Morten, & Sue, 1998; Sue & Sue, 2003; U.S. Census Bureau, 2002). One construct that has received significant attention from theorists and researchers is acculturation (see Atkinson et al., 1998; Sue & Sue, 2003). Acculturation was first defined by Redfield, Linton, and Herskovits (1936) as follows: "Acculturation comprehends those phenomena which result when groups of individuals sharing different cultures come into continuous first-hand contact, with subsequent changes in the original culture patterns of either or both groups" (p. 973). Three decades later, Graves (1967) coined the term *psychological acculturation* to describe the effects of acculturation at the individual level of study. Embedded in the concept of acculturation is the concept of *enculturation*. Herskovits (1948) defined enculturation as the process of socialization to (or retention of) the norms of one's indigenous culture. On the basis of this definition, Kim and Abreu (2001) explained that an adaptation process includes both adapting to the norms of the dominant culture and retaining the norms of the indigenous culture. They proposed that the term *acculturation* be used to describe the former process and the term *enculturation* be used to describe the latter process.

Consistent with this distinction, the current theory of acculturation (and enculturation) describes these two processes as occurring on two separate continua (Berry, 1990, 1994). According to Berry (1990, 1994), the process of adaptation can be categorized into four attitudes that are expressed in terms of the combined level of acculturation and enculturation: *integration*, *assimilation*, *separation*, and *marginalization*. Integration occurs when an individual becomes proficient in the culture of the dominant group (highly acculturated) while retaining proficiency in the indigenous culture (strongly enculturated); this status is also known as *biculturalism*. Assimilation occurs when an individual absorbs the culture

Sehee Hong, Department of Psychology, Ewha Womans University; Bryan S. K. Kim and Maren M. Wolfe, Department of Education, University of California, Santa Barbara. Correspondence concerning this article should be addressed to Sehee Hong, Department of Psychology, Ewha Womans University, 11-1 Daehyun-dong, Seodaemun-gu, Seoul, 120-750, Republic of Korea (e-mail: shong@ewha.ac.kr or bkim@education.ucsb.edu).

or Model

Wolfe, P. H. Yang,
psychometric
instrument of the

ated to the experi-
diverse and fastest
& Sue, 2003; U.S.
tion from theorists
03). Acculturation
Acculturation com-
ring different cul-
the original culture
(1967) coined the
n at the individual
t of *enculturation*.
to (or retention of)
1 and Abreu (2001)
as of the dominant
osed that the term
ulturation be used

and enculturation)
Berry, 1990, 1994).
orized into four at-
and enculturation:
n occurs when an
ighly acculturated)
ated); this status is
absorbs the culture

Kim, K. Kim and Maren
ra. Correspondence
ology, Ewha Womans
of Korea (e-mail:

of the dominant group (highly acculturated) while rejecting the indigenous culture (not enculturated). Separation occurs when an individual is not interested in learning the culture of the dominant group (not acculturated) and wants only to maintain and perpetuate the culture of origin (strongly enculturated). Finally, marginalization represents the attitude of an individual with no interest in maintaining or acquiring proficiency in any culture, dominant or indigenous (neither acculturated nor enculturated).

Regarding factors that define the constructs of acculturation and enculturation, Szapocznik, Scopetta, Kurtines, and Aranalde (1978) pointed out that cultural values represent one of two important components, with the dimension of behavior being the other component. More recently, Kim and Abreu (2001) noted that cultural values are one of four dimensions that constitute acculturation (and enculturation) processes—behaviors, knowledge, and cultural identity being the other three dimensions. Cultural values reflect relational style, person-nature relationships, beliefs about human nature, and time orientation (Szapocznik et al., 1978). P. B. Smith and Bond (1994) defined cultural values as “universalistic statements about what we think are desirable or attractive” (p. 52). Current theories on values acculturation/enculturation (i.e., acculturation and enculturation along the cultural values dimension) suggest that Asian Americans who are further removed from immigration (e.g., sixth-generation Asian Americans) will adhere to European American cultural values more strongly than will Asian Americans who are recent immigrants (Kim, Atkinson, & Umemoto, 2001). On the other hand, Asian Americans who are closer to immigration will adhere to Asian cultural values more strongly than their counterparts who are many generations removed from immigration. The theories also suggest that adherence to these values influences the ways in which Asian Americans behave, including how they manifest their psychological problems, express their emotions, and seek psychological help (Atkinson et al., 1998; Kim & Omizo, 2003; Sue & Sue, 2003). To facilitate further study of these ideas, the purpose of this study was to develop a psychometrically rigorous instrument, using the Rasch Model (Rasch, 1960), that could be used to assess values acculturation among Asian Americans (i.e., a measure of adherence to European American cultural values for Asian Americans).

Regarding existing measures of values acculturation for Asian Americans, Wolfe, Yang, Wong, and Atkinson (2001) developed the 18-item European American Values Scale for Asian Americans (EAVS-AA-18; Wolfe et al., 2001). The EAVS-AA-18 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*, 7 = *strongly agree*). Examples of items are “I think it is fine for an unmarried woman to have a child,” “The idea that one spouse does all the housework is outdated,” and “I’m confident in my ability to handle most situations.” Wolfe et al. selected items based on *t* tests of the mean scores on each of the 180 preliminary items; the 180 items were developed based on the theoretical and empirical literature on European American cultural values, including a sourcebook describing results of the World Values Survey (Inglehart, Basanez, & Moreno, 1993). Items that were scored significantly higher by European Americans than by first-generation Asian Americans were retained. This process led to the identification of 26 items that were more strongly endorsed by European Americans. However, Wolfe et al. determined that these items as a whole (i.e., as an instrument comprising all of the items) did not have adequate internal reliability (i.e., low coefficient alpha). Hence, the authors removed 8 items based on item score-total score statistics, resulting in the 18-item measure and an alpha coefficient of .69. In a further analysis, Wolfe et al. reported that an exploratory factor analysis did not converge after 200 iterations, and hence the dimensionality of this scale was unknown.

Although the development of the EAVS-AA-18 included appropriate procedures, the measure has several limitations. First, because the exploratory factor analysis did not yield an admissible result, it is not known whether the 18 items of the EAVS-AA-18 represent a homogenous construct. Second, because the items were chosen solely on the basis of *t* tests of group level scores between first-generation Asian Americans and European Americans of any generation, the functional characteristics of each item are not known. For in-

stance, it would be useful to know the extent to which each item is difficult to endorse. To illustrate, it could be reasoned that the EAVS-AA-18 item "A woman who is living alone should be able to have children" could be more difficult to endorse than another EAVS-AA-18 item "I think it is fine for an unmarried woman to have a child." The former item might lead a respondent to perceive that the woman's situation is riskier, in terms of not having child-rearing support, than is the latter situation, which does not invoke any contextual information regarding child rearing. It could be assumed that the latter situation may involve live-in relatives, partners, or friends. Third, it would be helpful to know whether or not the difficulty levels of the items reflect the full range of respondents' trait levels (i.e., degrees to which different respondents adhere to European American values). The EAVS-AA-18 items may be such that respondents with high adherence to U.S. values may not be fully distinguishable from someone with moderate adherence to the values. Fourth, it would be useful to know how well the 7-point scale captures the distinctions between each category of agreement. More specifically, it is unclear how well the 7-point scale is able to differentiate individuals who rate an item as a 2 (*moderately disagree*) from those who rate an item as a 3 (*mildly disagree*), or persons who rate an item as a 5 (*mildly agree*) from those who rate an item as a 6 (*moderately agree*). In addition to distinctions between levels of agreement, it is unclear whether the rating 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.

The successful development of a values measure of acculturation could complement an existing values measure of enculturation and allow researchers to study Berry's (1990, 1994) ideas regarding the combined effects of these two phenomena. Recently, the Asian Values Scale (AVS; Kim, Atkinson, & Yang, 1999), a measure of adherence to Asian cultural values for Asian Americans (i.e., values enculturation) that is a parallel version of the EAVS-AA-18, was revised based on the Rasch Model (Rasch, 1960). Although the AVS score has evidence of reliability and validity according to classical test theory, each of the 36 items had not been subjected to a more rigorous examination. For instance, like the EAVS-AA-18, because the items were chosen based solely on *t* tests of group level scores, the functional characteristics of each item were not known. The results of the Rasch analysis led to several important changes in the AVS. First, the 7-point scale was changed to a 4-point scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*) because it better represented participant responses. Second, 7 items that detracted from the AVS's construct homogeneity and 4 items that were found to be redundant were removed. The results showed that the remaining 25 items reflected the full range of person trait and item difficulty levels; this revised instrument was named the Asian Values Scale-Revised (AVS-R; Kim & Hong, 2004). Given the successful revision of the AVS, we attempted to apply the Rasch Model to revise the EAVS-AA-18 in the present study.

The need for psychometrically sound measures of values acculturation becomes apparent when reviewing the current research literature on values enculturation. In several studies on counseling process and outcome with Asian American clients and European American counselors, the construct of values enculturation was examined using the AVS. Kim and Atkinson (2002), on the basis of single sessions of counseling, found that Asian American clients with low adherence to Asian values rated their European American counselor to be more empathic than did clients with high adherence to Asian values. However, in another similar study, Kim, Li, and Liang (2002) found that Asian American clients with high adherence to Asian values perceived greater empathy and working alliance with their European American counselor than did clients with low adherence to Asian values. In a third study, Kim et al. (2003) found no relationship between Asian American client adherence to Asian values and session outcome with European American counselors. On the basis of these inconsistent findings, Kim et al. (2003) speculated that perhaps client adherence to U.S. values (i.e., values acculturation) might be affecting the relationship between client adherence to Asian values and session ratings. For instance, among Asian Americans who have

ult to endorse. To
ho is living alone
nother EAVS-AA-
er item might lead
not having child-
ntextual informa-
may involve live-
er or not the diffi-
ls (i.e., degrees to
AVS-AA-18 items
ot be fully distin-
it would be useful
category of agree-
differentiate indi-
ate an item as a 3
those who rate an
of agreement, it is
flects the halfway
together different

omplement an ex-
rry's (1990, 1994)
, the Asian Values
cultural values for
EAVS-AA-18, was
re has evidence of
tems had not been
A-18, because the
nal characteristics
important changes
ngly disagree, 2 =
participant responses.
14 items that were
; 25 items reflected
ent was named the
sful revision of the
n the present study.
becomes apparent

In several studies
ropean American
he AVS. Kim and
at Asian American
an counselor to be
wever, in another
ts with high adher-
ith their European
s. In a third study,
dherence to Asian
the basis of these
adherence to U.S.
ween client adher-
mericans who have

high adherence to Asian cultural values, their perceptions of European American counselors might be different if they had high rather than low adherence to U.S. values. Although now the more psychometrically rigorous AVS-R is available to examine this possibility, it is impossible to fully test this important hypothesis without having an equally good measure of values acculturation.

In sum, the goal of the present study was to revise the EAVS-AA-18 using a more rigorous and extant psychometric method, namely the Rasch Model (Rasch, 1960). In so doing, we sought to (a) investigate whether the 7-point scale is appropriate for the items, (b) modify the measure to increase the items' homogeneity, and (c) study the revised measure's items in terms of their relative distributions of the person trait level and item difficulty level estimates on a common logit scale.

METHOD

Participants

Respondents were 257 (112 male, 145 female) Asian American students at two community colleges ($n = 90$) and two universities ($n = 167$) in California. They ranged in age from 18 to 64 years, with a mean of 21.6 years ($SD = 4.5$). Seventy (27.2%) were seniors, 67 (26.1%) were sophomores, 63 (24.5%) were juniors, 53 (20.6%) were 1st-year students, 3 (1.2%) were graduate students, and 1 (.4%) did not report grade level. In terms of Asian ethnicity, there were 110 (42.8%) Chinese, 37 (14.4%) Korean, 29 (11.3%) Vietnamese, 26 (10.1%) Filipino, 22 (8.6%) Japanese, 22 (8.6%) other, 4 (1.6%) Asian Indian, 3 (1.2%) Cambodian, and 4 (1.6%) who did not report ethnicity. In terms of generation since immigration, 148 (57.6%) were first generation, 93 (36.2%) were second generation, 6 (2.3%) were fourth generation, 3 (1.2%) were third generation, 3 (1.2%) were fifth generation, 1 (.4%) was other, and 3 (1.2%) did not report generation status. (The preceding percentages do not total 100% due to rounding.)

Procedure

In the present study, we used the original data obtained by Wolfe et al. (2001) from college students at two universities and two community colleges in southern California for their development of the EAVS-AA-18. Wolfe et al. collected the data in three ways: (a) instrument distributed by the research team and completed in class; (b) instrument distributed by the research team, completed outside of the class, and collected by the course instructor; or (c) instrument distributed by the course instructor, completed outside of the class, and collected by the course instructor. For all three methods, the same instructions were read to student respondents.

Data Analysis

The basic Rasch Model (Rasch, 1960) is a dichotomous response model that 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; logit refers to log odd ratios, and the Rasch Model transforms ordinal data to logits in order to place items and persons on an interval-level scale. The Rasch Model assumes that all items 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 item difficulty parameter does not depend on the characteristics of the persons taking the measure and the person trait parameter does not depend on the items.

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). Assuming that item i is scored $x = 1, 2, \dots, m$ for $x = j$ the PCM can be expressed mathematically as

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

where the δ_{ij} term is called the item step difficulty associated with category j . In the PCM, the δ_{ij} parameter represents the relative difficulty of each step. This difficulty level can be viewed as varying levels of "hurdles" for people who are already in one category to overcome in order to jump to the next one. The higher the value of a particular δ_{ij} , the greater the difficulty level of a particular step relative to other steps within an item (Embretson & Reise, 2000). On the other hand, 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 δ_j parameter is estimated. The parameter δ_j , 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.

Although Likert scales can be modeled according to either a PCM or an RSM, the RSM was used for the present study. As explained above, in the PCM the step thresholds are allowed to vary for each item by introducing step parameters. Although the additional parameters can improve the fit of the data to the model, they usually reduce the interpretability of the model. Because steps are not uniform across items, score comparisons are problematic. Given that the purpose of scale construction is generally to develop measuring devices that are consistent across individuals, over-parameterizing should be avoided to account for idiosyncracies of the data (Piquero, MacIntosh, & Hickman, 2001). Another rationale for using the RSM is that, in the case under study here, there is no theoretical reason to expect that category usage by individuals in the samples would vary across items. Finally, note that for this study the Rasch analysis was undertaken using the WINSTEPS program with its default setting (Linacre & Wright, 1999).

RESULTS

Psychometric Properties of the EAVS-AA-18

Test of category use of the EAVS-AA-18. 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 that the test constructor intended (Roberts, 1994). Because it is always uncertain how individuals in a

sforms ordinal data
The Rasch Model
The Rasch Model has
assumption (Dinero &
The Rasch Model is its "spe-
parameter does not
person trait param-

ed categories. The
del (PCM; Masters,
as originally devel-
eps, in which partial
cess (Embretson &
M can be expressed

yj. In the PCM, the
level can be viewed
ory to overcome in
he greater the diffi-
mbretson & Reise,
tricts the step struc-
he RSM, a common
: also known as the
l distances between

an RSM, the RSM
step thresholds are
h the additional pa-
e the interpretability
risons are problem-
elop measuring de-
d be avoided to ac-
001). Another ratio-
o theoretical reason
ry across items. Fi-
ng the WINSTEPS

luating how respon-
ysis (Lopez, 1996).
manner that the test
ow individuals in a

sample use a rating scale, 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 scale category usage is to examine the average measure and threshold of each scale category. The average measures across scale categories represent the empirical averages of the measure (Rasch trait, ability, or endorsement score) that are modeled to produce the responses observed in each scale category (Linacre & Wright, 1998). Because observations in higher scale categories must be produced by higher measures, the average measures across scale categories must increase monotonically. The thresholds indicate the measures at which adjacent scale categories are equally probable and thus define the boundaries between the scale categories. Therefore, the thresholds should also increase monotonically.

The test of scale category use was performed using the EAVS-AA-18 items. The results showed that the average measure does not increase monotonically with the scale category label (with average measures of .15, .03, .06, .22, .44, .62, and .81 for Scale Categories 1 to 7, respectively). This result makes the meaning of the rating scale uncertain, and any measures derived under these circumstances are of doubtful utility (Linacre, 2002). Furthermore, threshold estimates are also not ordered (with logits -.46, -.07, -.43, .50, .13, and .32 for Scale Categories 2 to 7, respectively). Failure of threshold parameters to increase monotonically is known as "step disordering" (Linacre, 2002). Step disordering indicates a low probability of observance of certain scale categories. An examination of probability curves indicated that Scale Categories 3, 5, and 6 are never modal, indicating that at no point on the continuum are these scale categories ever the most likely scale category to be observed. That is, regardless of a person's trait level, the probability of choosing Scale Categories 3, 5, and 6 are never the most likely. This step disordering indicates that the EAVS-AA-18 scale may have categories that are poorly defined in the minds of the respondents (Linacre, 2002).

Dimensionality of the EAVS-AA-18. To study whether the EAVS-AA-18 scale supported a unidimensional structure, we computed item fit mean square (MNSQ) values using the RSM. MNSQ determines how well each item contributes to defining one common construct. According to R. M. Smith, Schumacker, and Bush (1995), item MNSQ values of about 1.0 are ideal, and values greater than 1.3 and less than 0.7 are unacceptable. 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). It should be noted, however, that these cutoff values are not absolute rules and can vary depending on the purpose for which the ratings are used (Karabatsos, 1997). Typically, two MNSQ statistics are used: *infit* and *outfit* statistics. Infit is more sensitive to misfitting responses to items near the person's ability (or endorsement) level, whereas outfit is sensitive to misfitting items that are farther away (Piquero et al., 2001). It is important to note that although infit and outfit statistics may not provide a complete dimensionality test, they nonetheless are related to and provide important information about dimensionality.

Among the original 18 items of the EAVS-AA-18, 3 items showed both infit and outfit statistics greater than 1.3, indicating a lack of construct homogeneity with other items. Three items showed both infit and outfit statistics less than 0.7, indicating redundancy with other items. Consequently, these results suggest that one third of the 18 items should be replaced.

Appropriateness of the EAVS-AA-18 difficulty level. As stated earlier, the Rasch Model estimates person trait and item difficulty parameters on a common metric, that is, logits. Using these parameters, it is possible to test whether the difficulty level of the EAVS-AA-18 items is appropriate for the sample in this study. If an instrument is appropriately targeted for the sample being tested, there should be sufficient overlap between the range of the person trait levels and that of item difficulty levels. Hence, we examined a distribution of person trait levels and item difficulty on a logit scale for the 18 items. The distribution showed that the logits for 8 items were at or below the trait level of the person with the lowest logits, thereby suggesting that these 8 items are too easy to agree with for the targeted sample

being tested. Because these items do not contribute much to measuring the targeted sample's trait level, they should be replaced with items with appropriate difficulty level.

In summary, the evaluation of the EAVS-AA-18 scale showed several problems in its psychometric properties. First, it seems that some categories are poorly defined in the minds of the respondents. Second, the EAVS-AA-18 scale did not support a unidimensional structure, given that 3 items were found to lack construct homogeneity. In addition, 3 items were found to be redundant with other items. Third, the difficulty level of the EAVS-AA-18 scale is not appropriately targeted for the sample being tested. Eight items were found to be too easy to agree with, and thus these items provide little, if any, additional significant information. Given these psychometric problems of the EAVS-AA-18 scale, we developed a revised scale based on data from the original 180 items in Wolfe et al. (2001).

Development of the Revised Scale

Identification of eligible items from the 180 Items. To identify possible items for inclusion in the revised version of the EAVS-AA-18, the 180 items from the original dataset were again submitted to *t* tests. Similar to the process of Wolfe et al. (2001), those items that were scored more highly by European American respondents than by first-generation Asian Americans were retained. However, instead of using an alpha level of .05 as was done in Wolfe et al., we used an alpha level of .10 so that more items could be identified. Using this more liberal alpha level allowed us to retain a sufficiently large set of items for possible inclusion in the revised instrument, while ensuring that these items still represented, to some degree, those value statements that distinguished first-generation Asian Americans from European Americans of any generation. The results yielded a total of 49 items, including the original 18 items.

Test of category use. First, we performed a Rasch rating scale analysis using data from the 49 items that were identified in the previous step. The results showed that the average measure did not increase with the category label (with average measures of .00, -.01, .05, .15, .34, .48, and .61 for Categories 1 to 7, respectively), suggesting that the rating scale categorization is not satisfactory. Furthermore, threshold estimates were not ordered (with logits -.42, -.11, .33, .40, .14, and .32 for Categories 2 to 7, respectively). This step disordering indicates a low probability of observance of certain categories. Also, an examination of probability curves indicated that Categories 2, 3, 5, and 6 were never modal.

On the basis of these results, we conducted a follow-up Rasch analysis to identify the optimal categorization. First, Category 2 was collapsed with Category 3, and Category 5 was collapsed with Category 6 of the scale to see if the threshold increased monotonically. Specifically, the original Category 1 was retained, but the original Categories of 2 and 3 were rescored as 2, the original Category of 4 as 3, the original Categories of 5 and 6 as 4, and the original Category of 7 as 5. This procedure produced the rescoring method of 1223445. However, this rescoring method failed to achieve step ordering. In addition, other scoring schemas (e.g., 112333) similarly failed to produce monotonically increasing thresholds. Hence, we decided to omit the neutral category by treating it as missing data. This decision to eliminate the neutral category is supported by several studies that have shown that a neutral category often does not work as intended (Andrich, 1982; Bock & Jones, 1968; Dubois & Burns, 1975). Thus, we tested the rating scale without a neutral category, 122X334, where X indicates the category that is treated as missing. When we used the scoring method of 122X334, the four category thresholds increased monotonically at approximately equal increments (with logits -1.68, .12, and 1.56 for Categories 2 to 4, respectively), suggesting support for this scoring method; the average measures were -.45, -.14, .42, and .86 for Categories 1 to 4, respectively. In further support of this scoring method, the corresponding probability curves as a function of item difficulty for each of the four response categories exhibited the desired appearance of a range of hills (see Figure 1). Hence, based on these results, we determined that the four-category scoring method (1 = *strong disagree*, 2 = *disagree*, 3 = *agree*,

he targeted sample's
ty level.
problems in its psy-
fined in the minds of
unidimensional struc-
ldition, 3 items were
EAVS-AA-18 scale
vere found to be too
significant informa-
developed a revised

tems for inclusion in
l dataset were again
ems that were scored
on Asian Americans
re in Wolfe et al., we
is more liberal alpha
clusion in the revised
ee, those value state-
an Americans of any
18 items.

ysis using data from
ved that the average
res of .00, -.01, .05,
that the rating scale
re not ordered (with
) . This step disorder-
o, an examination of
nodal.

lysis to identify the
, and Category 5 was
monotonically. Spe-
ries of 2 and 3 were
5 and 6 as 4, and the
method of 1223445.
dition, other scoring
ng thresholds. Hence,
his decision to elimi-
hown that a neutral
ies, 1968; Dubois &
, 122X334, where X
method of 122X334,
ly equal increments
ggesting support for
5 for Categories 1 to
ponding probability
gories exhibited the
ese results, we deter-
disagree, 3 = agree,

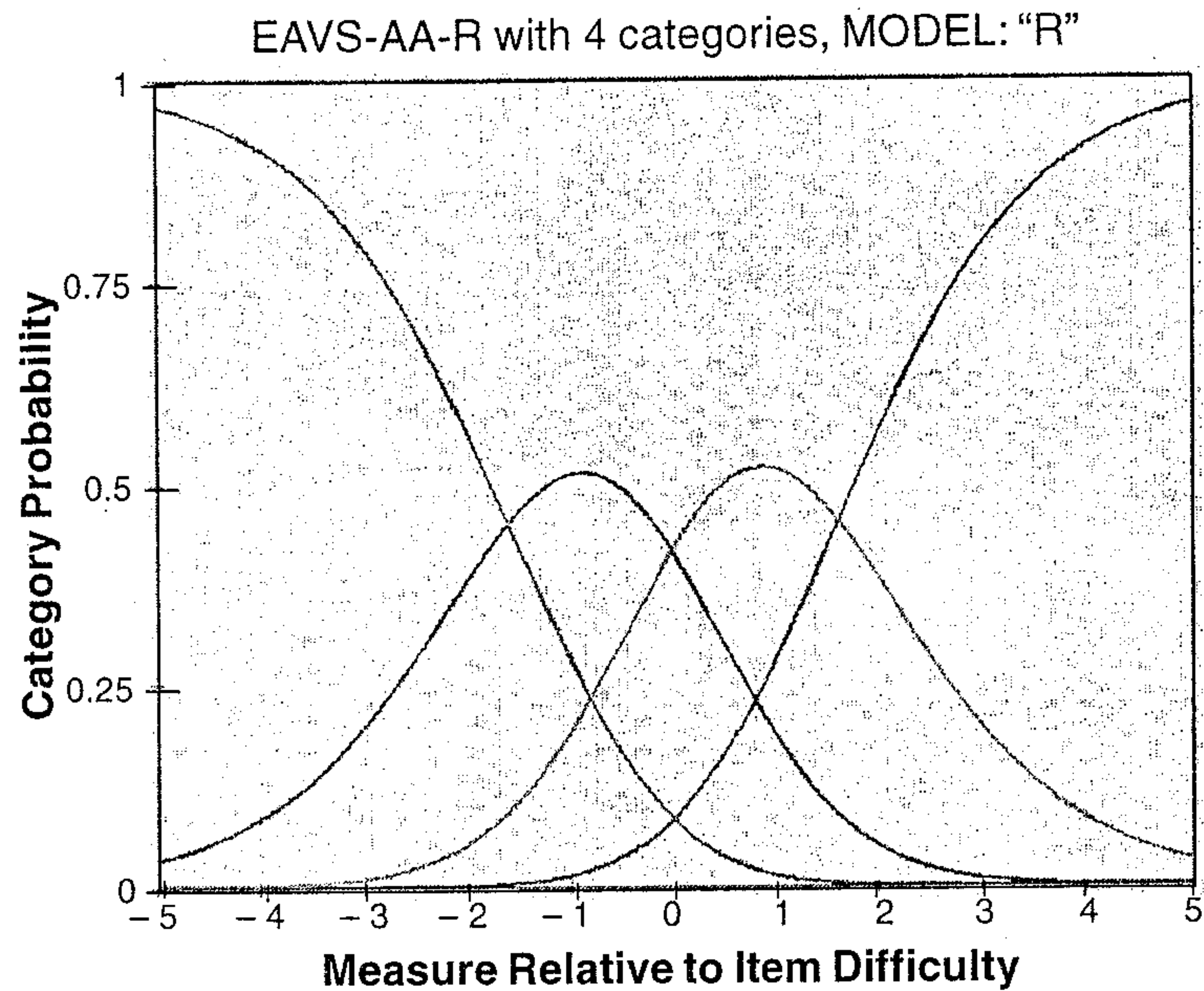


FIGURE 1

Response Functions for the Four Categories

Note. EAVS-AA-R = European American Values Scale for Asian Americans—Revised.

4 = *strongly agree*) is most appropriate for the revised scale. Note that the category thresholds and Figure 1 were developed based on the results obtained using the final 25-item revised scale that is described below.

Dimensionality. To achieve a unidimensional structure, MNSQ values of each item were evaluated using data from the initial pool of 49 items. According to their MNSQ values, 13 items were found to have either both infit and outfit statistics greater than 1.3 or both infit and outfit statistics less than 0.7. Because the responses for these items might be driven by other factors or the items might provide little information due to redundancy with other items, these 13 items were removed from the 49-item pool, leaving 36 items for possible inclusion in the final instrument.

Appropriateness of the difficulty level for the sample. A distribution of person trait level and item difficulty on a logit scale was developed using the 36 items. According to this distribution map, 11 items were found to be too easy to agree with for the targeted sample being tested. Because these items did not contribute much to measuring the targeted sample's trait level, these items were also removed, leaving a final set of 25 items.

In summary, based on response trends, the 7-point response anchor was reduced to a 4-point anchor. From the pool of 49 items, we deleted 24 items on the basis of unidimensionality, item difficulty, and respondent trait-level considerations, leaving a final set of 25 items. Note that the final set of items included 5 items that were originally included in the EAVS-AA-18. These 25 items constituted the revised instrument, which was named the European American Values Scale for Asian Americans—Revised (EAVS-AA-R).

Psychometric Evaluation of the EAVS-AA-R

To evaluate the psychometric properties of the EAVS-AA-R score, we conducted another RSM analysis 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, as mentioned previously, and there were nearly equal distances between categories (Figure 1). As shown in Table 1, no items except for Item 1 produced both infit and outfit values greater than 1.3 and less than 0.7. Because even the infit and outfit values of Item 1 were also close to 1.3, all of the items showed reasonable fit to the RSM. Next, we evaluated the appropriateness of the EAVS-AA-R's difficulty level for the present sample. Figure 2 displays a distribution person trait level on the left side of the axis and a distribution of item difficulty level on the right side of the axis. The coverage of the 25 items, which spans the entire range of a person's ability (or endorsement), indicates that this set of items is appropriately targeted for the sample being measured. The considerable overlap on both sides of the scale further indicates that the deleted items were superfluous items in terms of difficulty level.

TABLE 1
Parameter Values for the EAVS-AA-R's 25 Items (in Order of Item Difficulty)

Item Number and Statement	MNSQ		
	Difficulty	Infit	Outfit
47. Being humble is better than expressing feelings of pride.*	1.10	1.09	1.08
11. I prefer not to take on responsibility unless I must.*	0.99	0.87	0.86
21. It is OK to allow others to restrict one's sexual freedom.*	0.90	0.92	0.91
38. Monetary compensation is not very important for a job.*	0.79	0.84	0.83
30. It is OK for a woman to have a child without being in a permanent relationship.	0.67	0.91	0.91
29. I cannot approve of abortion just because the mother's health is at risk.*	0.45	1.00	0.99
3. Sometimes, it is necessary for the government to stifle individual development.*	0.36	0.90	0.90
44. Cheating on one's partner doesn't make a marriage unsuccessful.*	0.36	0.99	0.98
45. Greater emphasis on individual development is not a good thing.*	0.28	0.77	0.78
12. I do not like to serve as a model for others.*	0.24	0.80	0.81
23. A woman should not have a child unless she is in a long-term relationship.*	0.09	1.02	1.02
46. I have always enjoyed serving as a model for others.	0.05	0.93	0.93
28. Partners do not need to have similar values in order to have a successful marriage.*	0.00	1.17	1.23
9. Single women should not have children and raise them alone.*	-0.04	0.90	0.91
20. It is OK if work interferes with the rest of my life.*	-0.06	0.86	0.86
43. Luck determines the course of one's life.*	-0.30	1.28	1.30
42. A student does not always need to follow the teacher's instructions.	-0.34	1.12	1.14
27. The world would be a better place if each individual could maximize his or her development.	-0.54	1.05	1.04
37. Faithfulness is very important for a successful marriage.	-0.56	0.94	0.94
22. No one is entitled to complete sexual freedom without restriction.*	-0.58	1.29	1.43
4. You can do anything you put your mind to.	-0.62	0.98	0.99
48. Faithfulness is not important for a successful marriage.*	-0.64	0.91	0.90
25. I follow my supervisor's instructions even when I do not agree with them.*	-0.71	0.90	0.95
1. I think it is fine for an unmarried woman to have a child.	-0.92	1.40	1.43
36. Friends are very important.	-0.98	0.95	0.93

Note. EAVS-AA-R = European American Values Scale for Asian Americans-Revised; MNSQ = item fit mean square value. Asterisk indicates reverse-worded item.

categories were
by equal distances
Item 1 produced
even the infit and
reasonable fit to the
culty level for the
eft side of the axis
coverage of the 25
nt), indicates that
The considerable
were superfluous

Item Difficulty)

Item	Outfit
9	1.08
7	0.86
2	0.91
14	0.83
11	0.91
10	0.99
30	0.90
39	0.98
77	0.78
30	0.81
22	1.02
93	0.93
17	1.23
90	0.91
86	0.86
28	1.30
12	1.14
05	1.04
94	0.94
29	1.43
98	0.99
91	0.90
90	0.95
40	1.43
95	0.93

Revised; MNSQ = item

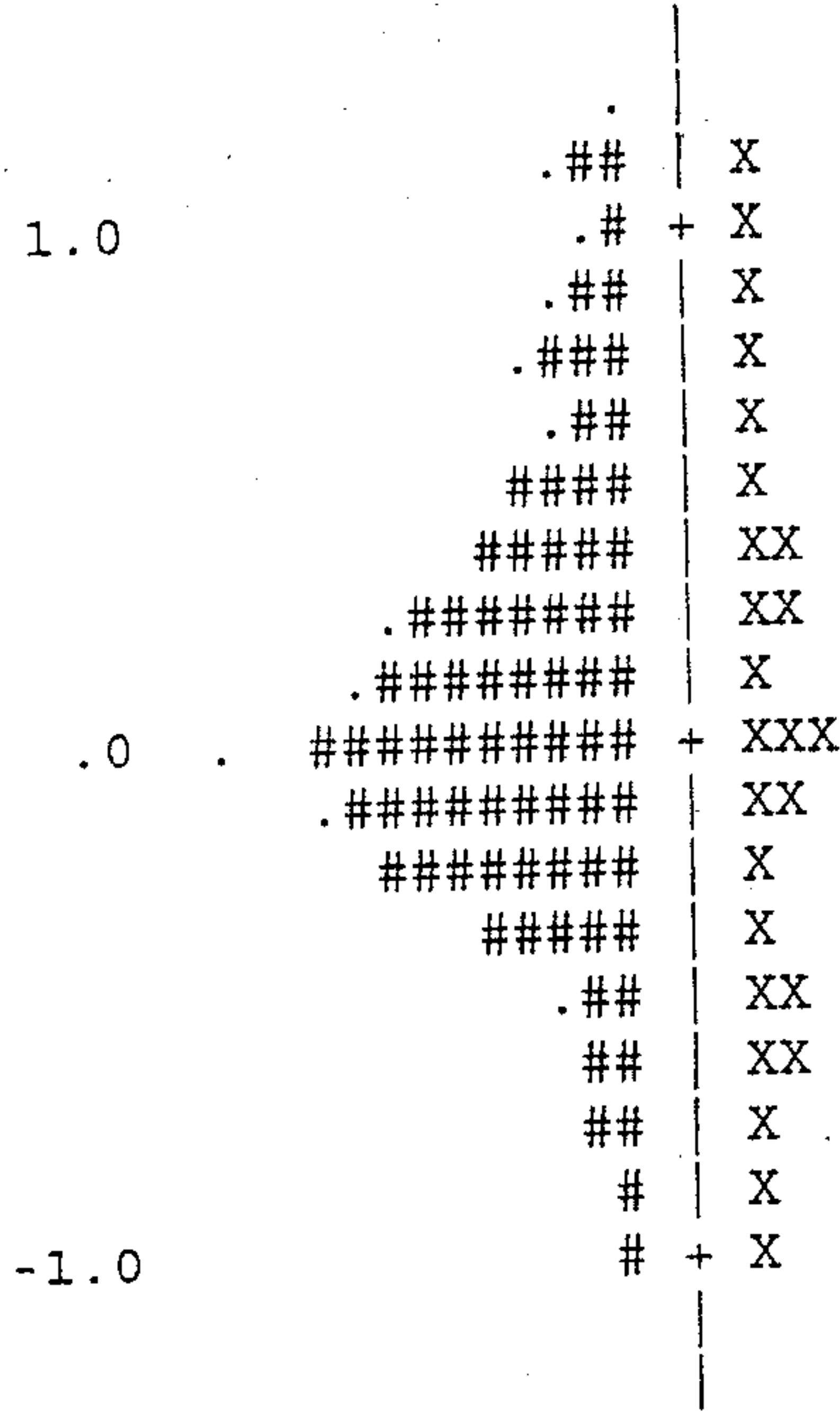


FIGURE 2

A Comparison of Person Trait Level and Item Difficulty Level for the EAVS-AA-R's 25 Items

Note. Each “#” represents two participants, and each “.” represents one person. Each “X” indicates one item. EAVS-AA-R = European American Values Scale for Asian Americans–Revised.

Finally, we computed the Rasch reliability indices. 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. We also computed the *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. Both indices range from 0 to 1, with values equal to or greater than .80 being regarded as acceptable (Fox & Jones, 1998). The person and item separation reliability indices of the EAVS-AA-R score were .78 and .98, respectively. In comparison, the coefficient alpha of the EAVS-AA-R score with the new 4-point scale was .77, with item-total correlations ranging from .02 to .61.

DISCUSSION

The original EAVS-AA-18 was designed to assess Asian Americans’ adherence to European American cultural values, an index of acculturation. As described earlier, despite a sufficient theoretical grounding and methodological sophistication in developing this instrument, the EAVS-AA has significant psychometric limitations. In the present study, we attempted to address these limitations on the basis of results from a psychometric analysis using the Rasch Model (Rasch, 1960). As a result, the 25-item EAVS-AA-R with a 4-point scale was established.

The finding that the original Category 4 (*neither agree nor disagree*) on the 7-point rating continuum was inadequate is important because it illustrates the problems that result from including a neutral category. It is often assumed that a neutral category reflects a midway

point on a Likert-type scale, but as we found in the present study, Category 4 might be unrelated to the *disagree-agree* continuum or, at least, poorly defined by respondents. Also, we deleted Categories 3 (*mildly disagree*) and 5 (*mildly agree*) because they were found to be inadequate in fully representing participant responses. Although these categories originally might have been used because it was assumed they would lead to a more precise index of agreement, the present findings suggest that they may lead instead to confusion on the part of the respondents regarding the definitions for these categories. 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* do not accurately represent an equal level of intensity (i.e., increase of thresholds at equal increments).

The second step in establishing the EAVS-AA-R involved examining the adequacy of the 18 items in the scale based on infit and outfit statistics (i.e., MNSQ values). Six items, or one third of the scale's items, either lacked construct homogeneity or were redundant with other items. One of the items that indicated lack of construct homogeneity was "Having the chance to achieve is not important to me in a job" (reverse worded). The present results indicate that this item represents different constructs in comparison with the construct assessed by the rest of the EAVS-AA-18 items. Among the redundant items was the following: "It is important to me to serve as a model for others." The results indicate that this item is too similar to other items and hence does not add new information. Furthermore, the results yielded 8 items that were too easy for respondents in the present sample to endorse and did not add to the variance explained by the scale. Therefore, we attempted to identify items that were more psychometrically valid using data from the original 180 items in Wolfe et al. (2001).

Using *t* tests, infit and outfit MNSQ values, and an examination of person trait-item difficulty map, we reduced the 180 items to 25 items, which then constituted the EAVS-AA-R. These items were found to represent a homogenous construct, without redundancy, and a full range of person trait and item difficulty levels. In addition, the scores on the 25 items were found to be internally consistent. These results indicate that the EAVS-AA-R score is reliable and adequately reflects individuals at the full range of values acculturation. As mentioned earlier, it is important to have an instrument that assesses both individuals with a high level of adherence to U.S. cultural values and a low level of adherence to U.S. values. The lack of sensitivity that allows assessment of individuals at either end of the trait level continuum could lead to either a ceiling or a floor effect and the resulting inability to use the instrument to detect the full variability in a population. Hence, these results provide evidence of the EAVS-AA-R score's validity.

The items in the EAVS-AA-R describe a range of situations related to U.S. cultural norms. An examination of these items provides interesting information about the various contexts in which one can observe U.S. cultural values that distinguish first-generation Asian Americans and general European Americans. The situations include child-rearing practices, marital status and behavior, autonomy, sexual freedom, career development, and friendship obligations. Hence, it could be speculated that Asian Americans who do not adhere to U.S. cultural values (i.e., low acculturated) might experience values conflicts across a wide range of settings.

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 EAVS-AA-R was developed using data from this population, the adequacy of the EAVS-AA-R with other Asian American populations should be examined before it is used with them. Second, because the data were obtained from a relatively small sample residing on the West Coast, 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. Fourth, the relatively small sample size for the RSM analyses may limit the external validity of the parameter estimates.

category 4 might be
y respondents. Also,
they were found to
se categories origi-
more precise index
to confusion on the
urthermore, the lack
e.g., between *mildly
gly agree*) suggests
rately represent an
s).

ng the adequacy of
) values). Six items,
/ or were redundant
ogeneity was "Having
ed). The present re-
rison with the con-
dondant items was
The results indicate
w information. Fur-
lents in the present
cale. Therefore, we
using data from the

son trait-item diffi-
ed the EAVS-AA-R.
dondancy, and a full
25 items were found
score is reliable and
mentioned earlier, it
gh level of adherence
ack of sensitivity that
could lead to either a
etect the full variabil-
A-R score's validity.
S. cultural norms. An
ous contexts in which
Asian Americans and
es, marital status and
obligations. Hence,
J.S. cultural values
range of settings.
lizability of the find-
the EAVS-AA-R was
A-R with other Asian
ond, because the data
t, the results may not
re present sample of
ient for those groups
d or not represented
limit the external va-

The results of the present study have implications for future research. First, the psychometric properties of scores from the EAVS-AA-R should be further examined in future research, including a study attempting to replicate the present findings with a new sample of Asian American college students. In addition, classical test theory might be used to further examine the EAVS-AA-R score's reliability and validity (both convergent and discriminant). For example, concurrent validity could be examined by comparing the EAVS-AA-R score with scores from other measures of acculturation (e.g., Suinn-Lew Asian Self-Identity Acculturation Scale; Suinn, Rickard-Figueroa, Lew, & Vigil, 1987). Second, in future research on Asian Americans' processes of adaptation to the norms of U.S. culture, the EAVS-AA-R might be used with the AVS-R to assess both values acculturation and enculturation, respectively. Such information could be helpful in identifying individuals at one of four adaptation statuses (integration, separation, assimilation, and marginalization) as theorized by Berry (1990, 1994). For instance, someone who scores high on both the EAVS-AA-R and the AVS-R could be classified as having an integration attitude, whereas someone who scores low on both measures could be classified as having a marginalization attitude. However, it would be important to first examine whether the use of the EAVS-AA-R and the AVS-R would indeed accurately identify the appropriate category for individual Asian Americans. Once this has been achieved, the instruments could be used to test the relationships among these categories, mental health status, and counseling process and outcome. Third, a more thorough investigation of the relationship between values and behavioral acculturation and enculturation should be conducted. Past research has indicated that values enculturation is inversely correlated with behavioral acculturation and that values enculturation tends to occur more slowly than behavioral acculturation (Kim et al., 1999). The availability of the EAVS-AA-R now allows researchers to examine the relationship between values acculturation and behavioral acculturation and also to examine whether the rate of values acculturation is slower than that of behavioral acculturation. Fourth, the successful application of the Rasch Model in developing the EAVS-AA-R illustrates the usefulness of this model. For this reason, researchers should consider using the Rasch Model in their work on instrument development. Fifth, future research should examine similarities and differences between various Asian ethnic groups on their levels of adherence to European American cultural values.

The present findings also have an implication for clinical practice. Counselors in practice may use the EAVS-AA-R in conjunction with the AVS to assess clients' levels of values acculturation and enculturation. Although normative data still need to be developed for both of these instruments, clinicians could use the items in each scale as a springboard for exploring various culture-related topics, including values conflict, acculturative and enculturative stresses, and worldview convergence and divergence.

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. 89-129). 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. (1990). Psychology of acculturation: Understanding individuals moving between cultures. In R. W. Brislin (Ed.), *Applied cross-cultural psychology* (pp. 232-253). Newbury Park, CA: Sage.
- Berry, J. W. (1994). Acculturation and psychological adaptation: An overview. In A. Bouvy, F. J. R. van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 129-141). Amsterdam: Swets & Zeitlinger.
- Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgement and choice*. San Francisco: Holden Day.

- 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.
- Graves, T. D. (1967). Psychological acculturation in a tri-ethnic community. *Southwestern Journal of Anthropology, 23*, 337-350.
- Herskovits, M. J. (1948). *Man and his works: The science of cultural anthropology*. New York: Knopf.
- Inglehart, R., Basanez, M., & Moreno, A. (1993). *Human values and beliefs: A cross-cultural sourcebook: Political, religious, sexual and economic norms in 43 countries from the 1990-1993 World Values Survey*. Ann Arbor: University of Michigan Press.
- 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., & Hong, S. (2004). A psychometric revision of the Asian Values Scale using the Rasch Model. *Measurement and Evaluation in Counseling and Development, 37*, 15-27.
- 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., & 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.
- 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.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.
- Piquero, A. R., MacIntosh, R., & Hickman, M. (2001). Applying Rasch modeling to the validity of a control balance scale. *Journal of Criminal Justice, 29*, 493-505.
- 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.
- Redfield, R., Linton, R., & Herskovits, M. J. (1936). Memorandum on the study of acculturation. *American Anthropologist, 56*, 973-1002.
- Roberts, J. (1994). Rating scale functioning. *Rasch Measurement Transactions, 8*, 386.
- 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.

varying item discrimi-
k response category in
Mahwah, NJ: Erlbaum.
logy research. *Journal*
Northwestern Journal of
ogy. New York: Knopf.
s-cultural sourcebook:
190-1993 World Values
y. *Journal of Outcome*
rrent instruments, and
nder (Eds.), *Handbook*
ge.
n cultural values, coun-
s. *Journal of Counseling*
the counseling process:
hologist, 29, 570-603.
e: Development, factor
342-352.
(2003). Counselor self-
and counseling process.
s Scale using the Rasch
. 15-27.
ient adherence to Asian
1 career counseling pro-
ward seeking profes-
nseling Psychologist,
Applied Measurement, 3,
MESA Press.
o-facet models (Version
ransactions, 10, 482-483.
a, 47, 149-174.
ling to the validity of a
ts. Chicago: Mesa Press.
or generality and validity
philosophy (pp. 58-94).
of acculturation. *Ameri-*
s, 8, 386.
nd perspectives. London:
in squares to evaluate fit
an Educational Research

- Sue, D. W., & Sue, D. (2003). *Counseling the culturally diverse: Theory and practice* (4th 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.
- Szapocznik, J., Scopetta, M. A., Kurtines, W., & Aranalde, M. A. (1978). Theory and measurement of acculturation. *Interamerican Journal of Psychology*, 12, 113-120.
- U.S. Census Bureau. (2002). *Census brief. The Asian population: 2000*. Retrieved April 3, 2002, from <http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf>
- Wolfe, M. M., Yang, P. H., Wong, E. C., & Atkinson, D. R. (2001). Design and development of the European American Values Scale for Asian Americans. *Cultural Diversity and Ethnic Minority Psychology*, 7, 274-283.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.