Reliability of Bidimensional Acculturation Scores: A Meta-Analysis

Que-Lam Huynh, Ryan T. Howell and Verónica Benet-Martínez

Journal of Cross-Cultural Psychology 2009; 40; 256
DOI: 10.1177/0022022108328919

Journal of Cross-Cultural Psychology
Volume 40 Number 2
March 2009 256-274

10.1177/0022022108328919
http://jccp.sagepub.com
hosted at
http://online.sagepub.com

Reliability of Bidimensional Acculturation Scores

A Meta-Analysis

Que-Lam Huynh
University of California, Riverside
Ryan T. Howell
San Francisco State University
Verónica Benet-Martínez
University of California, Riverside

Understanding score reliability is a necessary step in examining the validity of acculturation instruments. Thus, the authors evaluate the aggregate reliability of three multigroup, bidimensional acculturation instruments: General Ethnicity Questionnaire–Abridged, Stephenson Multigroup Acculturation Scale, and Vancouver Index of Acculturation. Reliability generalization techniques are used to analyze 51 internal consistency estimates and 6 sample characteristics for these instruments. Overall, reliability estimates for all three instruments appear to be robust. However, the ranges of reliability estimates across diverse samples vary greatly, which has implications for the interpretation of substantive outcomes and acculturation instrument selection. In addition, variability in reliability estimates is associated with scale length, gender, and ethnic composition of sample. Implications for acculturation research and measurement are discussed.

Keywords: *acculturation, meta-analysis, reliability generalization, General Ethnicity Questionnaire, Stephenson Multigroup Acculturation Scale, Vancouver Index of Acculturation*

Acculturation at the individual, psychological level results when there is continuous, prolonged contact between two culturally distinct groups, and individuals within the groups must adapt to the new cultural contact situation (Berry, 2003). Within recent years, researchers have widely embraced the bidimensional or bilinear model of acculturation, which proposes that individuals can have independent orientations toward their heritage culture and the host or dominant culture (i.e., independent cultural orientations; see Berry, 2003; Cuellar, Arnold, & Maldonado, 1995; Ryder, Alden, & Paulhus, 2000; Stephenson,

Authors' Note: Portions of this paper were presented at the 20th annual convention of the Asian American Psychological Association, August 2005. We are indebted to Angela-MinhTu D. Nguyen, Robert Rosenthal, and M. Robin DiMatteo for their invaluable feedback on previous drafts of this article. Address correspondence to Que-Lam Huynh, Department of Psychology, University of California, 900 University Ave., Riverside, CA 92521; e-mail: que-lam.huynh@email.ucr.edu.

2000; Tsai, Ying, & Lee, 2000). This model replaced the older one-dimensional models, which were based on the perception that as individuals spend more time in a second culture, they simultaneously must become more oriented toward that culture and relinquish their heritage culture (e.g., Cuellar, Harris, & Jasso, 1980; Suinn, Rickard-Figueroa, Lew, & Vigil, 1987). Along with this theoretical revision, there has been a shift in measurement toward bidimensionality (Zane & Mak, 2003). In this study, we examined and compared three widely-used, multigroup, bidimensional measures of acculturation: the General Ethnicity Questionnaire–Abridged (GEQ; Tsai et al., 2000), the Stephenson Multigroup Acculturation Scale (SMAS; Stephenson, 2000), and the Vancouver Index of Acculturation (VIA; Ryder et al., 2000). Specifically, we conducted a meta-analysis of their score reliability (i.e., performed a reliability generalization or RG; see Henson & Thompson, 2002; Vacha-Haase, 1998, for details about RG methodology) and explored moderators of score reliability across the samples in which they have been administered.

We focus on the reliability of these instruments because it is essential to identify the factors that contribute to variability in score reliability before examining the validity of such scores (Crocker & Algina, 1986; Nunnally & Bernstein, 1994). We also focus on measurement in general because measurement dictates the research done and consequently the findings in any given area of research, including acculturation research. For example, although bidimensional theories of acculturation have existed since at least the 1970s (e.g., see Berry, 1979), unidimensional measures of acculturation limited researchers to comparisons between assimilation (i.e., high dominant culture orientation and low nondominant culture orientation) versus separation (i.e., low dominant culture orientation and high nondominant culture orientation) strategies during the past few decades. Until recently, researchers could not examine all four acculturation strategies because the bicultural or integration strategy (i.e., high on both cultural orientations) was confounded with the marginalization strategy (i.e., low on both cultural orientations) when unidimensional measures were used (Nguyen & Benet-Martínez, 2007). Moreover, researchers could not examine the independent contribution of each cultural orientation to important outcomes, such as psychological well-being, distress, school or work performance, or sociocultural adjustment.

Even when scale development and item selection accurately reflect current theoretical perspectives (i.e., the researcher measures independent cultural orientations), the reliability of a scale places an upper limit on all validity coefficients and important outcome variables. A perfectly unreliable instrument yields a zero correlation between scores on that scale and any outcome (John & Benet-Martínez, 2000). Therefore, low score reliability can attenuate correlations between cultural orientations and important outcome variables (e.g., using a measure with score reliability of .60, the upper limit of the validity coefficient would be the square root of .60, which is approximately .77, when the outcome variable is measured with perfect reliability). Thus, without understanding aggregate score reliability, researchers may misinterpret important results in acculturation research, and this in turn may influence public policy and the distribution of resources for outreach and treatment programs for ethnic minorities (Chun, Organista, & Marín, 2003). Because score reliability is a necessary condition of validity, we must understand the score reliability of bidimensional acculturation measures before examining their validity.

Score Reliability

In its broadest sense, reliability is the extent to which measurements are repeatable or stable (Crocker & Algina, 1986; Nunnally & Bernstein, 1994). Thus, reliability is a property of the scores, not of tests. Furthermore, a given instrument can yield scores that are reliable for one sample and other scores that are unreliable for another sample (Rowley, 1976). Because score reliability is a characteristic of data, it is essential that researchers not only understand the average score reliability typically produced by a scale but also understand the influence of study characteristics and participants on score reliability.

Reporting practices. Reliability plays an important role in the measurement and understanding of psychological variables, yet characteristics of reliability are not well understood (Li, Rosenthal, & Rubin, 1996). It may be this lack of understanding that has led to the current trend of underreporting reliability indices in the published literature. For example, although it has been suggested that researchers report reliability estimates for each unique measurement (Wilkinson & the APA Task Force on Statistical Inference, 1999), reviews of reporting practices have shown that reliability reporting varies across entire journal volumes (Kieffer, Reese, & Thompson, 2001; Meier & Davis, 1990; Whittington, 1998) and psychological instruments (Vacha-Haase, Henson, & Caruso, 2002). These reviews suggest that authors in the areas of psychology examined tend to either (a) give no reliability information or (b) make a reference to reliability coefficients from past studies, thereby inferring that the data at hand must be as reliable as the data referenced.

Reliability generalization. RG is a meta-analytic technique that allows researchers to examine reported score reliability (Henson & Thompson, 2002; Vacha-Haase, 1998), and it is an extension of validity generalization (see Hunter & Schmidt, 1990; Schmidt & Hunter, 1977). RG is used to (a) determine the average reliability of scores generated by a given test and describe the variability in score reliability and (b) examine the relationship between score reliability and participant and sample characteristics (e.g., mean sample age, ethnic or gender composition). As with other meta-analytic methods, RG allows researchers to document characteristics by which score reliability may vary systematically.

Acculturation Measures of Interest

The multigroup, bidimensional measures of acculturation of interest in this study are the GEQ (Tsai et al., 2000), SMAS (Stephenson, 2000), and VIA (Ryder et al., 2000). First, these measures were chosen because they are in line with current conceptualizations of acculturation as a bidimensional process. Second, they were developed for use with all ethnic groups, whereas most other bidimensional measures of acculturation were intended for specific ethnic groups (e.g., Acculturation Rating Scale for Mexican Americans–Revised by Cuellar et al., 1995). Third, these measures have gained considerable popularity, and they are used by many researchers to measure acculturation in various groups, across disparate locales. Because of their widespread use (e.g., a search in Social Sciences Citation Index in January 2007 shows that together, these instruments have been cited almost 120 times since their publication in 2000) and potential applicability across different groups, it is important to examine how instrument characteristics (e.g., length) and sample characteristics (e.g., age, gender, ethnic composition) are associated with variations in score reliability of these measures. Therefore, in this study we assess the reliability generalization of these three instruments using RG.

Method

Acculturation Measures

Although researchers have referred to the two groups involved in the acculturation process by various names, we use Berry's (2003) terminology in the following descriptions because it is the most widespread in acculturation research. *Nondominant culture* refers to the heritage or ethnic group, or to the heritage or ethnic cultural orientation, and *dominant culture* refers to the mainstream or host group, or to the mainstream or host cultural orientation.

General Ethnicity Questionnaire–Abridged (GEQ). The GEQ (Tsai et al., 2000) is a 77-item questionnaire that assesses the extent to which respondents participate in and identify with the nondominant and dominant cultures. Each cultural orientation subscale has 38 items (plus 1 item about bilingualism). The items are identical in wording, except for the culture referenced. These items measure six domains of acculturation: language use and proficiency, social affiliation, cultural participation, cultural pride, cultural exposure, and media (dominant culture scale) or food (nondominant culture scale) preference. The GEQ items are rated on a 5-point Likert-type scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Stephenson Multigroup Acculturation Scale (SMAS). The SMAS (Stephenson, 2000) is a 32-item questionnaire that assesses the extent to which respondents are immersed in the nondominant and dominant cultures. The dominant culture scale has 15 items, and the nondominant culture scale has 17 items. The wording of SMAS items varies by scale, but each scale measures the same domains: language, interaction, media, and food. Within each domain, the items reflect knowledge, behaviors, and attitudes (e.g., language knowledge, language behavior, and language attitude). The SMAS items are rated on a 4-point Likert-type scale, ranging from 1 (*false*) to 4 (*true*).

Vancouver Index of Acculturation (VIA). The VIA (Ryder et al., 2000) is a 20-item questionnaire that assesses the extent to which respondents participate in and identify with the nondominant and dominant cultures. Each cultural orientation subscale has 10 items, which are identical in wording except for the culture referenced. These items assess three domains of acculturation: values, social relationships, and adherence to traditions. The VIA items are rated on a 9-point Likert-type scale, ranging from 1 (*strongly disagree*) to 9 (*strongly agree*).

Literature Search and Data Collection

To identify the studies in which at least one of these three measures of acculturation was administered, we performed a literature search using PsycINFO. The search terms used were “General Ethnicity Questionnaire,” “Stephenson Multigroup Acculturation Scale,” and “Vancouver Index of Acculturation” as keywords. In addition, we used PsycINFO and the Social Sciences Citation Index in a forward search to identify all of the articles that have cited Tsai et al. (2000), Stephenson (2000), and Ryder et al. (2000). For the present study, the search was limited to dissertations and articles published in peer-reviewed journals in English prior to January 2007 (when data were gathered and coded). Finally, we sent e-mail requests to (a) members of two psychology listservers, the Asian American Psychological Association and the APA’s Society for the Psychological Study of Ethnic Minority Issues, and (b) the first authors of the GEQ, SMAS, and VIA validation studies to obtain unpublished data and manuscripts to be included in the RG.

Inclusion and Exclusion Criteria

After obtaining the articles and unpublished manuscripts, we confirmed that each article/manuscript used at least one subscale of the GEQ, SMAS, or VIA. We eliminated false-positive citations and articles/manuscripts that used significantly different versions of one of the measures (e.g., Beaupré & Hess, 2006, used a shortened, French-language version of the VIA; Nguyen, Huynh, & Lonergan-Garwick, 2007, used a shortened version of the GEQ), because these would decrease generalizability of our findings. One additional article/manuscript (Toyama, 2005) was eliminated because it reported unidimensional scores for the bidimensional SMAS. Finally, we manually reviewed the reference lists of the remaining articles/manuscripts to identify those not captured in the original search. All articles/manuscripts were coded and entered into SPSS 12.0 by the first author. After excluding irrelevant articles/manuscripts, we identified a total of 39 empirical articles/manuscripts using at least one subscale of the GEQ, SMAS, or VIA and scoring it bidimensionally. Because some articles/manuscripts reported data from more than one independent sample, we coded 66 unique measurements for the nondominant and dominant culture scores.

Data Reduction and Coding

Of the 66 unique VIA, SMAS, and GEQ measurement opportunities, 9 (13.6%) made no reference to score reliability. Five (7.6%) made references to reliability estimates from previous data, specifically to the reliability estimates from the validation study of the instrument administered. Finally, 52 (78.8%) were coupled with a reliability estimate based on data gathered for that particular study.

We were able to obtain reliability estimates from some of the authors who did not originally report score reliability in their publications. Thus, we had a total of 59 unique reliability estimates. However, further inspection revealed that 8 of these 59 studies (13.6%) provided reliability coefficients reported in previous research on the same sample. These were excluded from analyses, leaving 51 reports of an independent administration of at least one subscale of the VIA, SMAS, or GEQ. These 51 unique measurements of nondominant or dominant culture score reliability were used in this RG (see the appendix).

Coding of samples. Of these 51 unique samples, the majority (39 samples or 76.5% of all samples) were composed of U.S. ethnic minorities or immigrants, and the remaining 12 samples (23.5%) were from various countries around the world. Of the 12 samples from outside the United States, 7 were from Canada. For most samples, the group under study was a numerical minority in its country (an exception included European Americans who were used as a comparison group in some studies and thus given only the dominant culture subscale). Thus, we coded ethnicity as follows: African descent (3.9%), European descent (25.5%), Latin descent (9.8%), indigenous groups within country of study (2.0%), Chinese descent (17.6%), and other Asian (but non-Chinese) descent (23.5%). All of these groups included immigrants as well as ethnic minorities in the countries of study (e.g., the "African descent" group included first-generation African immigrants in the United States and U.S.-born African Americans). For the articles/manuscripts that did not report reliability coefficients separately for each ethnic group, we coded ethnicity as diverse (17.6%). Because there was only one study with an indigenous sample, this study was coded as the "miscellaneous" group in multiple regression analyses with effect coding. Age was coded as mean age reported for the sample, and gender was coded as percentage of females in the sample. Finally, generation status was coded as proportion of participants in the sample who were born outside the country of study.

Predictor and Criterion Variables in Data Analysis

Internal consistency reliability estimates, particularly coefficient alpha, are the most commonly reported estimates of reliability. Because alpha was reported for all 51 samples, it was the estimate of reliability in this RG. Based on classical test theory, scale length is a likely predictor of score reliability, with longer scales producing scores with higher internal consistency (Crocker & Algina, 1986). In addition, important demographic variables that make a sample of participants more or less heterogeneous than the validation sample may also predict score reliability, including participant age, generational status, gender, ethnicity, and country of study (United States or Canada, from which the majority of samples came). Therefore, in our regression analyses, alpha was the criterion variable, and scale length and demographic variables with at least small to moderate associations ($r \geq .20$) with the reliability coefficients were predictor variables.

Random Versus Fixed Effects Analysis

Random and fixed effects analyses differ in three important ways. First, in a random effects analysis, the unit of analysis is the study, whereas in a fixed effects analysis, the unit of analysis is the participants within the study. Thus, in a fixed effects analysis, the reliability estimates for each study are weighted by sample size. As a result, random effects analyses are more conservative and considerably less powerful than fixed effects analyses. Second, in a random effects analysis, each study contributes equally to the overall reliability estimate, whereas in a fixed effects analysis, studies with more participants contribute proportionally more than studies with fewer participants. Third, random effects analysis allows for generalization to the population of possible studies, whereas fixed effects analysis allows for generalization only to the population of participants who could have been in the studies included in the RG. Despite the tradeoff for lower statistical power, we deemed it appropriate given our sample size to use the random effects model to gain broader generalizability of our findings.

Results

Characteristics of Samples Analyzed

Of the 51 unique samples included in this RG, there was a wide range of sample sizes, from 30 to more than 600 respondents. The typical sample included about 100 respondents (M = 166, Mdn = 105, SD = 148). Most samples (k = 39) were from the United States, and the typical sample included college-age respondents (M = 22.47, Mdn = 20.41, SD = 5.73). Of the 50 samples for which we had gender composition information, nearly half (42%) had approximately equal (45% to 55% of each group) proportions of male and female respondents (proportion female: M = 58.07, Mdn = 54.25, SD = 21.98). Finally, the typical sample included about one third foreign-born participants (proportion born outside country of study: M = 36.53, Mdn = 29.00, SD = 33.33).

Characteristics and Homogeneity of Score Reliability

Table 1 shows the estimates of central tendency and dispersion of coefficient alpha in the 51 samples included in this RG analysis (note that in some samples, only one culture orientation scale was administered). The mean and median reported reliability coefficients for the three acculturation measures have been at or above the conventional cutoff value of .80 for research purposes (Nunnally & Bernstein, 1994), except for the coefficients for the dominant culture scale of the SMAS (which has a mean and median reliability greater than .70). The descriptive statistics reported in Table 1 show that the VIA, SMAS, and GEQ usually generate acceptably reliable scores, but they cannot always be counted on to do so from study to study. In fact, 10 (23.8%) nondominant culture orientation reliability estimates and 11 (23.4%) dominant culture orientation reliability estimates were below the conventional cutoff for most research purposes (.80), and 30 (71.4%) nondominant culture orientation reliability estimates and 37 (78.7%) dominant culture orientation reliability estimates were below the conventional cutoff for most clinical purposes (.90; Nunnally & Bernstein, 1994). For the correlation and regression analyses reported below, we excluded the one case in which dominant culture orientation alpha was .39 (see the appendix), because this could be considered an outlier—no other reliability estimate was that low. The descriptive statistics reported in Table 1 for dominant culture orientation reliability estimates did not change significantly when the outlier was removed, but it was not included in correlation and regression analyses based on recommendations by Cohen, Cohen, West, and Aiken (2003).

Correlates of Reliability

First, we tested the relationship between scale length and score reliability estimates using bivariate correlations (see Table 2) and analysis of variance. Scale length was correlated moderately and positively with dominant culture score reliability (r = .46, p = .001), such that longer dominant culture scales tended to produce more reliable scores. To explore this relationship,

Table 1
Descriptive Statistics for Reliability Coefficients and Demographic Variables (Excluding Ethnicity)

						% Below Cutoffs	
Variable	*k*	*M*	*Mdn*	*SD*	Range	Research	Clinical
Nondominant culture orientation: alpha for all measures	42	.84	.87	.09	.63–.97	23.8	71.4
Vancouver Index of Acculturation	14	.83	.85	.08	.66–.92	28.6	78.6
Stephenson Multigroup Acculturation Scale	12	.82	.85	.11	.63–.97	25.0	75.0
General Ethnicity Questionnaire (Abridged)	16	.86	.88	.07	.64–.93	18.8	62.5
Dominant culture orientation: alpha for all measures	47	.83	.85	.09	.39–.94	23.4	78.7
Vancouver Index of Acculturation	13	.83	.84	.06	.70–.89	15.4	100.0
Stephenson Multigroup Acculturation Scale	11	.74	.77	.14	.39–.90	72.7	90.9
General Ethnicity Questionnaire (Abridged)	23	.87	.88	.04	.77–.94	4.3	60.9
Predictor variables							
Average age	51	22.47	20.41	5.73	15.70–40.80	—	—
Proportion female	50	58.07	54.25	21.98	0.00–100.00	—	—
Proportion born outside country of study	39	36.53	29.00	33.33	0.00–100.00	—	—

Note: Total *N* of 51 samples yielded 8,491 respondents. In some samples, only one culture orientation scale was administered; thus, the sample sizes for nondominant and dominant culture orientation are less than 51. The conventional cutoff for research purposes is $\alpha = .80$, and the conventional cutoff for clinical purposes is $\alpha = .90$ (Nunnally & Bernstein, 1994).

we conducted a univariate analysis of variance comparing reliability estimates separately for nondominant and dominant scores for the GEQ, SMAS, and VIA. There were no significant differences among the nondominant culture reliability estimates for the three acculturation measures, $F(2, 39) = 0.78$, $p = .47$, $\eta^2 = .04$. However, there were large and significant differences among the dominant culture reliability estimates, $F(2, 44) = 11.25$, $p = .0001$, $\eta^2 = .34$.

Table 2
Correlations Between Reliability Coefficients and Predictor Variables

Variable	*n*	*r*	*p*
Average age			
Nondominant culture orientation	42	.05	.753
Dominant culture orientation	46	.09	.538
Proportion female			
Nondominant culture orientation	41	–.05	.775
Dominant culture orientation	45	–.40	.006
Country of study (United States vs. Canada)			
Nondominant culture orientation	37	.009	.957
Dominant culture orientation	41	.05	.974
Proportion born outside country of study			
Nondominant culture orientation	32	.15	.406
Dominant culture orientation	37	.05	.784
African descent			
Nondominant culture orientation	42	.18	.251
Dominant culture orientation	46	.17	.273
European descent			
Nondominant culture orientation	42	–.01	.950
Dominant culture orientation	46	–.05	.722
Latin descent			
Nondominant culture orientation	42	–.35	.024
Dominant culture orientation	46	–.55	.00007
Asian (non-Chinese) descent			
Nondominant culture orientation	42	–.14	.815
Dominant culture orientation	46	–.18	.237
Chinese descent			
Nondominant culture orientation	42	.21	.182
Dominant culture orientation	46	.52	.0002
Ethnically diverse sample			
Nondominant culture orientation	42	.11	.508
Dominant culture orientation	46	.09	.547
Scale length			
Nondominant culture orientation	42	.17	.283
Dominant culture orientation	46	.46	.001

Post hoc analyses revealed that the GEQ yielded marginally more reliable dominant culture scores than the VIA and significantly more reliable scores than the SMAS (Table 1). VIA and SMAS reliability estimates for the dominant culture scores were not significantly different from each other, $p = .17$.

Second, we tested the relationship between sample-specific variables (age, generation status, gender, and country of study) and score reliability estimates using bivariate correlations (see Table 2). Of these sample-specific variables, only gender and dominant culture score reliability were moderately and negatively correlated to each other ($r = -.40$, $p = .02$). Studies with a greater proportion of women tended to produce less reliable scores on the dominant culture scale. In terms of ethnicity, dominant culture score reliability was negatively associated with Latino ethnicity ($r = -.55$, $p = .00007$), but it was positively

Table 3
Regression Analyses Predicting Reliability Coefficients

Variable	Constant	*b*	*SE b*	β	*p*	Pearson *r*
Nondominant Culture reliability coefficient ($k = 42$)[a]						
Model	.84		.02		<.001	
Latin descent		–.08	.04	–.32	.04	–.35
Chinese descent		.04	.03	.18	.23	.21
Dominant Culture reliability coefficient ($k = 46$)[b]						
Step 1: $R^2 = .36$, $p = .0001$	.87		.026		<.001	
Proportion female		–.001	.0003	–.38	.004	–.40
Scale length		.002	.001	.44	.001	.46
Step 2: $\Delta R^2 = .310$, $p < .001$	.83		.02		<.001	
Proportion female		–.0004	.0003	–.12	.25	–.40
Scale length		.002	.001	.31	.02	.46
Latin descent		–.09	.02	–.49	.00001	–.55
Chinese descent		.06	.02	.39	.0003	.52

Note: For both regression analyses, the indigenous sample ($N = 1$) was the miscellaneous group. We only entered the moderately and/or significantly correlated predictors into each regression.
[a]Total variance explained in reliability coefficients for nondominant culture scores: $R^2 = .14$; $F(2, 39) = 3.112$, $p = .06$. [b]Total variance explained in reliability coefficients for dominant culture scores: $R^2 = .67$; $F(4, 40) = 19.85$, $p < .001$.

associated with Chinese ethnicity ($r = .52$, $p = .0002$). Also, nondominant culture score reliability was negatively associated with Latino ethnicity ($r = –.35$, $p = .024$). Studies with a greater proportion of respondents of Latin descent tended to produce less reliable scores on the dominant and nondominant culture scales, whereas studies with a greater proportion of respondents of Chinese descent tended to produce more reliable scores on the dominant culture scale. Note that ethnicity variables were coded dichotomously (e.g., Chinese vs. not Chinese), so correlations between ethnicity variables and alphas reported in Table 2 are Pearson point-biserial correlations.

Predicting Future Reliability

To predict future score reliability for these instruments, we regressed reliability coefficients onto the predictor variables with at least small to moderate associations ($r \geq .20$) with the reliability coefficients as reported in Table 2, even if these correlations were not significant. These regression models can help test users to estimate the reliability coefficient expected for a particular sample. In addition, these models can help to inform sampling decisions to improve that reliability coefficient when needed.

Table 3 summarizes results from the multiple regression analyses. In the first regression analysis, reliability coefficients for the nondominant culture orientation scale of all three acculturation measures were regressed onto the two variables with moderate correlations with score reliability: Latino and Chinese ethnicity (no other sample characteristics affect the reliability of the nondominant culture orientation scales). The constant in these regression analyses, .84 (the grand mean), is the reliability coefficient expected in a future study in which the nondominant culture scale from the VIA, SMAS, or GEQ is administered regardless of the ethnic composition of the sample. However, increasing the proportion of individuals of Latin descent can decrease alpha significantly, but increasing the proportion of individuals of Chinese descent does not significantly change alpha.

Next, reliability coefficients for the dominant culture orientation scale of all three acculturation measures were regressed onto gender and scale length (centered around 10 items, the number of items for each subscale in the shortest instrument, the VIA) in the first step and onto the three ethnicity variables with moderate correlations with score reliability (Latino, and Chinese ethnicity) in the second step (see Table 3). The constant, .83, is the reliability coefficient expected in a future study in which the dominant culture scale from the VIA, SMAS, or GEQ is administered, and this is the grand mean of the reliability coefficients for dominant culture subscales at mean levels of the predictors included in the equation. After ethnicity variables are introduced into the equation, the proportion of women is no longer a significant predictor of alpha, but scale length remains a significant predictor of alpha, such that increasing the scale length can increase alpha. Latino ethnicity predicts a medium to large, significant decrease and Chinese ethnicity predicts a medium, significant increase in alpha even after controlling for gender and scale length.

Discussion

On average, the GEQ, SMAS, and VIA have yielded alphas above .80 on both the nondominant and dominant culture scales across many different samples. Although these average score reliability estimates would meet conventional cutoffs for research purposes (.80), they would not meet conventional cutoffs for clinical settings (.90; Nunnally & Bernstein, 1994). Moreover, the range of score reliability estimates was high, with nondominant reliability estimates ranging from .63 to .97 and dominant reliability estimates ranging from .67 to .94 (not including the outlier of .39). In research and clinical settings, such a wide range of score reliability has serious consequences for validity coefficients and decision making. Although acculturation assessment tools such as these are typically used in research settings and would be adequate for such settings, the GEQ, SMAS, and VIA should be used with caution in situations where personal and social costs are considered more important or significant.

In addition to characterizing score reliability, we also examined factors relating to variation in score reliability using regression analyses. In RG studies, it is important to determine the factors that relate to score reliability, and it is equally important to examine those factors that do not contribute to score reliability. Researchers and test users can exercise caution when using the measures under conditions known to be related to lower score reliability. Conversely, researchers and test users can use the measures with more confidence under conditions known to be related to higher score reliability. In our study, scale length, gender, and ethnicity were all related to score reliability estimates, but there were different patterns of these relations for nondominant and dominant culture scales.

Factors Related to Score Reliability

Scale length. There was no substantial difference among scale lengths on reliability estimates for nondominant culture scores, but longer dominant culture scales tended to produce more reliable scores. Although typically, having more items in a scale results in higher score reliability, this is not necessarily the case when these items have little content overlap with each other. Given the complex and multidimensional nature of acculturation and the vast array of different nondominant groups undergoing acculturation (Berry, 2003), it is perhaps not surprising that for nondominant culture orientation, having more items is not related to having higher score reliability. As compared with the VIA, the longer instruments (SMAS and GEQ) measure more specific domains of acculturation (e.g., cultural pride, media use), which may represent different facets or subconstructs under the rubric of culture orientation. Because nondominant cultural orientation may vary depending on the domain and the culture (e.g., group status, history, etc.), having more items that measure different domains of acculturation would likely not increase score reliability. On the other hand, orientation to dominant culture may be less varied in content than orientation toward nondominant culture across the different domains of acculturation. For example, most of these samples were taken from the United States, so being oriented to the dominant American culture may look similar across groups in different geographical locations within the United States.

Gender composition. There was no substantial difference between male and female samples on reliability estimates for nondominant culture scores, but samples with more women tended to produce less reliable dominant culture scores. It is not entirely clear why dominant culture scores should be less reliable for women. One possible explanation is that the dominant culture items were less internally consistent for women because of gender differences in the acculturation process. For example, it is possible that for men, all of the items measure orientation to dominant culture, whereas for women, they measure orientation to dominant culture as well as opportunity to interact with members of the dominant group outside the home, opportunity to learn the dominant language, and so on. Thus, for women, the dominant culture items might be measuring several different (but still related) constructs, making them less internally consistent. Regardless of the reason behind this finding, the dominant culture scales of these instruments should be used with more caution for women.

Ethnic composition. Reliability of nondominant and dominant culture scores was related to sample ethnicity. Both subscales for all three instruments yielded less reliable scores for respondents of Latin descent. Latinos were underrepresented in all standardization samples of the GEQ, SMAS, and VIA, which suggests that the questionnaire items did not hang together as well for Latinos, which may partially account for the lower reliability of scores for this group. On the other hand, the dominant culture subscale yielded more reliable scores for those of Chinese descent. Respondents of Chinese descent were well represented in the GEQ and VIA standardization samples, so the items may hang together best for members of this group, which may account for their overall higher reliability estimates.

In summary, for nondominant culture score reliability, having more participants of Latin descent was related to a small to moderate decrease in score reliability, whereas having more participants of Chinese descent was related to a moderate increase in score reliability.

For dominant culture score reliability, having more participants who are female or of Latin descent also was related to moderate decreases in score reliability, whereas having more participants of Chinese descent or a longer scale was related to moderate increases in score reliability. Thus, for both nondominant and dominant culture scales of the GEQ, SMAS, and VIA, Chinese and Latino ethnicities were related consistently to changes in score reliability. This suggests that although these instruments yielded adequately reliable scores for both groups, more focused inquiries are needed concerning the performance of these multigroup acculturation instruments for Latinos.

Limitations

There are several important limitations of our analyses. First, we made every effort to obtain unpublished data and score reliability estimates when they were not reported in published articles, but as in every meta-analysis, there is the possibility of a file-drawer problem. Fail-safe *N* analyses indicated that the mean population reliability of nondominant culture orientation scores would be less than .80 (the commonly accepted threshold in research settings) if there were 23 file-drawer studies with an unweighted average reliability estimate of .728 (see Howell & Shields, 2008, for details about computing the fail-safe *N* for RG studies). For dominant culture orientation score, fail-safe *N* analyses indicated that the mean population reliability would be less than .80 if there were 20 file-drawer studies with an unweighted average reliability estimate of .728. For these reasons, it is imperative that more researchers report score reliability for nondominant and dominant culture orientation scales of the GEQ, SMAS, and VIA for us to be confident that the population reliability coefficients are above .80.

Given that more than one fourth of research articles and manuscripts obtained did not accurately report reliability estimates, we must reiterate that reliability is a property of scores, not of tests. Wilkinson and APA Task Force on Statistical Inference (1999) reminded researchers that "a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees. Thus, authors should provide reliability coefficients of the scores for the data being analyzed even when the focus of their research is not psychometric" (p. 596).

Another limitation is the small sample size and overrepresentation of U.S. samples, which contributes to lack of power and limits our ability to generalize results to other countries. We could only conduct analyses looking at the contribution of country of study for the United States and Canada (although the sample size for Canada was quite small, $k = 7$). Given the increased diversity and intergroup contact in many countries attributable to globalization, more acculturation research is needed in countries other than the United States. Moreover, it was not possible to conduct our analyses separately for the GEQ, SMAS, and VIA because of the small sample size for each instrument. Although we attempted to account for scale length differences in score reliability, scale length is confounded with the scale itself. Therefore, other characteristics such as item wording or measured content may have contributed to the wide range of score reliability estimates reported in the literature; for example, the range of alphas for nondominant culture orientation subscales was from .63 (unreliable scores) to .97 (very reliable scores). However, given the important similarities between these measures (all were developed for potential use with many different ethnic groups, and all are bidimensional), we deemed it appropriate to combine these measures for analyses. Furthermore, by aggregating them in our analyses, we were able to examine some common correlates of score reliability among the three instruments.

Implications

Despite these limitations, this is an important first-ever attempt to characterize and predict bidimensional acculturation score reliability. The bidimensional acculturation measures we analyzed were instrumental in moving the field forward in terms of measurement. They have (at least temporarily) settled the debate about unidimensional versus bidimensional acculturation measures. Now, given their wide use and popularity, we have added to the research literature by examining the reliability of the scores that they yield.

Specifically, our regression equations can be used by researchers and other test users to predict future score reliability for a sample. For example, the constant in the regression for nondominant culture orientation (Table 3) indicates that the expected alpha is .84 for a sample with an average number of Latinos or persons of Chinese descent. Adding betas allows us to predict that the alphas would be .76 (95% confidence interval = .69-.83) for a sample of Latinos and .88 (95% confidence interval = .82-.94) for a sample with persons of Chinese descent. These estimates allow researchers to understand how large of a difference in score reliability they can expect between samples of different ethnic groups. Although these sample or participant characteristics predict variation in score reliability, the constants in both equations were above .80, the conventional cutoff for research purposes.

In summary, we evaluated three multigroup, bidimensional acculturation instruments using RG, a meta-analytic method for characterizing score reliability. For the 51 unique samples analyzed, reliability estimates of the GEQ, SMAS, and VIA appear to be robust and were above conventional cutoffs for research purposes, although the ranges of reliability estimates were wide. Using bivariate correlations, we found that variability in the reliability estimates was associated with scale length, gender, and ethnic composition of the samples, but the pattern of association was different for nondominant and dominant culture orientations. Our multiple regression findings suggest that test users can be confident in the reliability of the GEQ, SMAS, and VIA scores, especially when research is done on groups in the United States, but there are some conditions under which these instruments should be used with caution. In addition, test users can use our equations to calculate the predicted reliability of the GEQ, SMAS, and VIA scores given their sample characteristics. Finally, our study underscores the need for researchers to accurately report score reliability in their publications, even when their research is not focused on the psychometric properties of one of these acculturation instruments.

Appendix
Reliability Estimates and Basic Demographics for Each Sample Included in Meta-Analysis

		Alpha							% of Each Ethnic Group					
Reference	Scale	N-D	D	Country	% Born Outside Country	*N*	M_{age}	% Female	African	European	Latin	Indigenous	Chinese	Other Asian
Amer (2005)	VIA	0.88	0.82	U.S.	35	609	29.3	60	0.0	100.0	0.0	0.0	0.0	0.0
Amer (2005)	VIA	0.8	0.84	U.S.	—	76	30.0	999	0.0	100.0	0.0	0.0	0.0	0.0
Bowen (2002)	SMAS	0.82	0.81	U.S.	100.0	35	38.6	63	0.0	0.0	100.0	0.0	0.0	0.0
Bowen (2002)	SMAS	0.63	0.67	U.S.	100.0	36	29.4	92	0.0	0.0	100.0	0.0	0.0	0.0
Cheah & Nelson (2004)	VIA	0.86	—	Canada	0.0	69	20.6	61	0.0	0.0	0.0	100.0	0.0	0.0
David & Okazaki (2006)	VIA	0.86	0.86	U.S.	37.0	603	28.9	66	0.0	0.0	0.0	0.0	0.0	100.0
D'urso (2004)	SMAS	0.94	0.87	U.S.	49.0	109	20.9	63	—	—	—	—	—	—
Hall et al. (2005)	SMAS	0.83	—	U.S.	—	399	21.5	0	0.0	100.0	0.0	0.0	0.0	0.0
Hall et al. (2005)	SMAS	0.67	—	U.S.	—	222	20.7	0	0.0	0.0	0.0	0.0	100.0	0.0
Hall et al. (2005)	SMAS	0.81	—	U.S.	—	127	22.3	0	0.0	0.0	0.0	0.0	0.0	100.0
Kennedy et al. (2005)	VIA	0.68	0.75	Canada	11.0	459	19.7	80	0.0	100.0	0.0	0.0	0.0	0.0
Kennedy et al. (2005)	VIA	0.77	0.88	Canada	65.0	574	19.3	75	0.0	0.0	0.0	0.0	100.0	0.0
Kennedy et al. (2005)	VIA	0.88	0.81	Canada	21.0	102	19.5	73	0.0	0.0	0.0	0.0	0.0	100.0
Lyubansky & Eidelson (2005)	GEQ	0.85	0.91	U.S.	—	100	40.5	71	100.0	0.0	0.0	0.0	0.0	0.0
Miville & Constantine (2006)	SMAS	0.66	0.73	U.S.	—	162	19.6	64	0.0	0.0	100.0	0.0	0.0	0.0
Rayle & Meyers (2004)	SMAS	0.89	0.78	U.S.	—	286	16.3	49	0.0	100.0	0.0	0.0	0.0	0.0
Rayle & Meyers (2004)	SMAS	0.86	0.79	U.S.	—	176	16.2	53	68.0	0.0	16.0	2.0	0.0	0.0
Raymundo (2005)	GEQ	0.87	0.9	U.S.	10.0	49	21.6	100	100.0	0.0	0.0	0.0	0.0	0.0
Raymundo (2005)	GEQ	0.79	0.82	U.S.	19.0	68	22.5	100	0.0	0.0	100.0	0.0	0.0	0.0
Raymundo (2005)	GEQ	0.64	0.77	U.S.	27.0	74	21.1	100	0.0	0.0	0.0	0.0	0.0	100.0
Ryder et al. (2000), Study 3	VIA	0.91	0.87	Canada	16.0	140	20.2	71	—	—	—	—	—	—
Ryder et al. (2000), Study 3	VIA	0.91	0.89	Canada	61.0	204	19.8	69	0.0	0.0	0.0	0.0	100.0	0.0
Ryder et al. (2000), Study 3	VIA	0.92	0.85	Canada	51.0	70	20.0	77	0.0	0.0	0.0	0.0	0.0	100.0
Stephenson (2000), Study 2	SMAS	0.97	0.9	U.S.	47.0	436	30.0	70	44.0	29.0	19.0	0.0	0.0	0.0
Stephenson (2000), Study 3	SMAS	0.94	0.75	U.S.	14.0	208	22.8	79	5.0	73.0	15.0	0.0	0.0	0.0
Swagler & Jome (2005)	VIA	0.84	0.8	Taiwan	100.0	125	34.4	36	0.8	91.0	2.4	0.0	0.0	0.0
Tsai (2001)	GEQ	0.88	0.81	U.S.	70.0	46	20.0	61	0.0	0.0	0.0	0.0	0.0	100.0
Tsai et al. (2000)	GEQ	0.92	0.92	U.S.	65.0	353	20.2	51	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2002[a])	GEQ	—	0.88	U.S.	0.0	44	20.6	46	0.0	100.0	0.0	0.0	0.0	0.0

(continued)

Appendix (continued)

		Alpha							% of Each Ethnic Group					
Reference	Scale	N-D	D	Country	% Born Outside Country	*N*	M_{age}	% Female	African	European	Latin	Indigenous	Chinese	Other Asian
Tsai et al. (2002[a])	GEQ	0.87	0.84	U.S.	71.0	49	20.1	59	0.0	0.0	0.0	0.0	0.0	100.0
Tsai et al. (2002[b]), Study 1	GEQ	0.9	0.9	U.S.	62.0	144	20.2	50	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2002[b]), Study 1	GEQ	—	0.81	U.S.	0.0	170	20.8	50	0.0	100.0	0.0	0.0	0.0	0.0
Tsai et al. (2002[b]), Study 2	GEQ	0.88	0.83	U.S.	72.0	50	20.1	52	0.0	0.0	0.0	0.0	0.0	100.0
Tsai et al. (2002[b]), Study 2	GEQ	—	0.87	U.S.	0.0	48	20.7	48	0.0	100.0	0.0	0.0	0.0	0.0
Tsai et al. (2004), Study 1	GEQ	0.91	0.94	U.S.	67.0	30	22.2	53	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2004), Study 1	GEQ	—	0.94	U.S.	0.0	30	20.8	53	0.0	100.0	0.0	0.0	0.0	0.0
Tsai et al. (2004), Study 2	GEQ	0.92	0.91	U.S.	64.0	96	20.4	50	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2004), Study 2	GEQ	—	0.87	U.S.	0.0	100	20.8	50	0.0	100.0	0.0	0.0	0.0	0.0
Tsai et al. (2006), Study 1	GEQ	0.77	0.85	U.S.	—	196	19.8	54	—	—	—	—	c	c
Tsai et al. (2006), Study 1	GEQ	—	0.86	U.S.	—	201	20.0	52	0.0	100.0	0.0	0.0	0.0	0.0
Tsai et al. (2006), Study 2	GEQ	0.9	0.88	U.S.	—	81	19.7	48	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2006), Study 2	GEQ	0.86	0.92	HK	—	96	20.2	49	0.0	0.0	0.0	0.0	100.0	0.0
Tsai et al. (2006), Study 2	GEQ	—	0.86	U.S.	—	79	20.0	54	0.0	100.0	0.0	0.0	0.0	0.0
Tsoh et al. (2003)	GEQ	0.93	0.92	U.S.	97.0	199	40.8	21	0.0	0.0	0.0	0.0	100.0	0.0
Varte & Zokaitluangi (2006)	VIA	0.82	0.7	India	0.0	260	20.1	54	0.0	0.0	0.0	0.0	0.0	100.0
Varte & Zokaitluangi (2006)	VIA	0.78	0.81	India	0.0	105	21.2	51	0.0	0.0	0.0	0.0	0.0	100.0
Varte & Zokaitluangi (2006)	VIA	0.66	0.88	India	0.0	170	19.1	42	0.0	0.0	0.0	0.0	0.0	100.0
Vergara (2006)	SMAS	—	0.67	U.S.	17.0	76	19.2	84	0.0	0.0	100.0	0.0	0.0	0.0
Vergara (2006)	SMAS	—	0.39	U.S.	2.9.0	104	19.0	69	0.0	100.0	0.0	0.0	0.0	0.0
Winograd (2005)	SMAS	0.88	0.77	U.S.	29.0	102	18.6	81	37.0	0.0	51.0	0.0	0.0	0.0
Wong (2001)	GEQ	0.88	0.88	U.S.	45.0	144	15.7	51	0.0	0.0	0.0	0.0	c	c

Note. Percentages of each ethnic group for some studies do not add to 100% because of missing information. Those samples were coded as “ethnically diverse” in the regression analyses. A dash in a cell indicates that data were not obtained or not reported. Alpha N-D = alpha for nondominant culture orientation scores; Alpha D = alpha for dominant culture orientation scores; African = participants of African descent; European = participants of European descent; Latin = participants of Latin descent; Indigenous = participants who belonged to indigenous groups within country of study; Chinese = participants of Chinese descent; Other Asian = participants of other Asian (but non-Chinese) descent; HK = Hong Kong; VIA = Vancouver Index of Acculturation; SMAS = Stephenson Multigroup Acculturation Scale; GEQ = General Ethnicity Questionnaire–Abridged.

[a]Tsai, Chentsova-Dutton, Freire-Bebeau, and Przymus (2002). [b]Tsai, Mortensen, Wong, and Hess (2002). [c]These samples included Chinese and other Asian (non-Chinese) groups, and we could not obtain sample size for each group separately. They were coded as “ethnically diverse” in the regression analyses.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Amer, M. M. (2005). Arab American mental health in the post September 11 era: Acculturation, stress, and coping (Doctoral dissertation, University of Toledo, 2005). *Dissertation Abstracts International*, *66*, 1974.

Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin*, *32*, 16-26.

Berry, J. W. (1979). Research in multicultural societies: Implications of cross-cultural methods. *Journal of Cross-Cultural Psychology*, *10*, 415-434.

Berry, J. W. (2003). Conceptual approaches to acculturation. In K. M. Chun, P. B. Organista, & G. Marín (Eds.), *Acculturation: Advances in theory, measurement, and applied research* (pp. 17-37). Washington, DC: American Psychological Association.

*Bowen, N. A. (2002). Satisfaction with life of refugees and immigrants (Doctoral dissertation, University of Texas, Austin, 2002). *Dissertation Abstracts International*, *64*, 3515.

*Cheah, C. S. L., & Nelson, L. J. (2004). The role of acculturation in the emerging adulthood of aboriginal college students. *International Journal of Behavioral Development*, *28*, 495-507.

Chun, K. M., Organista, P. B., & Marín G. (Eds.). (2003). *Acculturation: Advances in theory, measurement, and applied research.* Washington, DC: American Psychological Association.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican Americans–II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Sciences*, *17*, 275-304.

Cuellar, I., Harris, L. C., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical populations. *Hispanic Journal of Behavioral Sciences*, *2*, 199-217.

*David, E. J. R., & Okazaki, S. (2006). The Colonial Mentality Scale (CMS) for Filipino Americans: Scale construction and psychological implications. *Journal of Counseling Psychology*, *53*, 241-252.

*D'urso, N. (2004). Effect of context on recall of self-referent and group-referent words in individuals from individualistic or collectivistic cultures (Doctoral dissertation, Hofstra University, 2004). *Dissertation Abstracts International*, *64*, 5199.

*Hall, G. C. N., Teten, A. L., DeGarmo, D. S., Sue, S., & Stephens, K. A. (2005). Ethnicity, culture, and sexual aggression: Risk and protective factors. *Journal of Consulting and Clinical Psychology*, *73*, 830-840.

Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies. *Measurement and Evaluation in Counseling and Development*, *35*, 113-126.

Howell, R. T., & Shields, A. L. (2008). The file drawer problem in reliability generalizations: A strategy to compute a fail-safe *N* with reliability coefficients. *Educational and Psychological Measurement*, *68*, 120-128.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

John, O. P., & Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339-369). New York: Cambridge University Press.

*Kennedy, M. A., Parhar, K. K., Samra, J., & Gorzalka, B. (2005). Suicide ideation in different generations of immigrants. *Canadian Journal of Psychiatry*, *50*, 353-356.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *Journal of Experimental Education*, *69*, 280-309.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*, 98-107.

*Lyubansky, M., & Eidelson, R. J. (2005). Revisiting Du Bois: The relationship between African American double consciousness and beliefs about racial and national group experiences. *Journal of Black Psychology*, *31*, 3-26.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology*, *37*, 113-115.

*Miville, M. L., & Constantine, M. G. (2006). Sociocultural predictors of psychological help-seeking attitudes and behavior among Mexican American college students. *Cultural Diversity and Ethnic Minority Psychology*, *12*, 420-432.

Nguyen, A.-M. D., & Benet-Martínez, V. (2007). Biculturalism unpacked: Components, individual differences, measurement, and outcomes. *Social and Personality Psychology Compass*, *1*, 101-114.

Nguyen, A.-M. D., Huynh, Q.-L., & Lonergan-Garwick, J. (2007). The role of acculturation in the mentoring-career satisfaction model for Asian/Pacific Islander American university faculty. *Cultural Diversity and Ethnic Minority Psychology*, *13*, 295-303.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

*Rayle, A. D., & Myers, J. E. (2004). Counseling adolescents toward wellness: The roles of ethnic identity, acculturation, and mattering. *Professional School Counseling*, *8*, 81-90.

*Raymundo, M. M. (2005). *Body dissatisfaction and disordered eating among minority women: The roles of acculturation and standard comparisons.* Unpublished master's thesis, University of Houston, TX.

Rowley, G. L. (1976). The reliability of observational measures. *American Educational Research Journal*, *13*, 51-59.

*Ryder, A. G., Alden, L. E., & Paulhus, D. L. (2000). Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology*, *79*, 49-65.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529-540.

*Stephenson, M. (2000). Development and validation of the Stephenson Multigroup Acculturation Scale (SMAS). *Psychological Assessment*, *12*, 77-88.

Suinn, R. M., Rickard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn-Lew Asian Self-Identity Acculturation Scale (SL-ASIA). *Educational and Psychological Measurement*, *47*, 401-407.

*Swagler, M. A., & Jome, L. R. M. (2005). The effects of personality and acculturation on the adjustment of North American sojourners in Taiwan. *Journal of Counseling Psychology*, *52*, 527-536.

Toyama, K. M. (2005). *Self construal of Japanese-born women in the US: Examining egalitarian sex-role attitudes, self-esteem, acculturation and independence/interdependence*. Unpublished doctoral dissertation, New York University.

*Tsai, J. L. (2001). Cultural orientation of Hmong young adults. *Journal of Human Behavior in the Social Environment*, *3*, 99-114.

*Tsai, J. L., Chentsova-Dutton, Y., Freire-Bebeau, L., & Przymus, D. E. (2002). Emotional expression and physiology in European Americans and Hmong Americans. *Emotion*, *2*, 380-397.

*Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*, 288-307.

*Tsai, J. L., Mortensen, H., Wong, Y., & Hess, D. (2002). What does "Being American" mean? A comparison of Asian American and European American young adults. *Cultural Diversity and Ethnic Minority Psychology*, *8*, 257-273.

*Tsai, J. L., Simeonova, D. I., & Watanabe, J. T. (2004). Somatic and social: Chinese Americans talk about emotion. *Personality and Social Psychology Bulletin*, *30*, 1226-1238.

*Tsai, J. L., Ying, Y.-W., & Lee, P. A. (2000). The meaning of "being Chinese" and "being American": Variation among Chinese American young adults. *Journal of Cross-Cultural Psychology*, *31*, 302-331.

*Tsoh, J. Y., Lam, J. N., Delucchi, K. L., & Hall, S. M. (2003). Smoking and depression in Chinese Americans. *American Journal of the Medical Sciences*, *326*, 187-191.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6-20.

Vacha-Haase, T., Henson, R., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, *62*, 562-569.

*Varte, C. L., & Zokaitluangi, L. (2006). Aspects of identity and symptoms of adjustment related problems in the acculturating Mizo students. *Social Science International*, *22*, 111-132.

*Vergara, D. (2006). *Culture, parental attachment, and trait anxiety: A comparison of Latino- and Caucasian-Americans.* Manuscript submitted for publication.

Whittington, D. (1998). How well do researchers report their measures? An evaluation of measurement in published educational research. *Educational and Psychological Measurement*, *58*, 21-37.
Wilkinson, L., & the American Psychological Association Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.
*Winograd, G. (2005). Predictors of counseling expectations among students in an opportunity program (Doctoral dissertation, The City University of New York, 2005). *Dissertation Abstracts International*, *66*, 898.
*Wong, S. L. (2001). Depression level in inner-city Asian American adolescents: The contributions of cultural orientation and interpersonal relationships. *Journal of Human Behavior in the Social Environment*, *3*, 49-64.
Zane, N., & Mak, W. (2003). Major approaches to the measurement of acculturation among ethnic minority populations: A content analysis and an alternative empirical strategy. In K. M. Chun, P. B. Organista, & G. Marín (Eds.), *Acculturation: Advances in theory, measurement, and applied research* (pp. 39-60). Washington, DC: American Psychological Association.

Que-Lam Huynh, MA, is a social/personality psychology doctoral candidate at the University of California, Riverside. Her research interests include acculturation/enculturation theory and measurement, biculturalism, multiple identity negotiation, bicultural identity development, and cultural essentialism. Her major research focus is on identifying the underlying processes and commonalities among biculturals of different American ethnic groups as they negotiate conflicting value systems, norms, and dual identities.

Ryan T. Howell, PhD, is an assistant professor of psychology at San Francisco State University. He received his PhD in personality psychology from the University of California, Riverside (2005). His research attempts to determine effects of purchasing choices on psychological need satisfaction and well-being. He is also interested in using meta-analytic techniques to examine the variance in reliability and validity across alcohol screening measures.

Verónica Benet-Martínez, PhD, is an associate professor of psychology at the University of California, Riverside, where she heads the Culture and Personality Laboratory. She holds a PhD in social/personality psychology from the University of California, Davis, and was a postdoctoral researcher at the University of California, Berkeley. Veronica's main research interests are bicultural/multicultural identity and the interplay of culture and personality processes. She has been the IACCP regional representative from North America since 2005.