

Implications of Test Revisions for Assessment With Asian Americans

Sumie Okazaki
University of Wisconsin—Madison

Stanley Sue
University of California, Davis

There are serious gaps in knowledge with respect to the use of standardized assessment instruments such as the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; D. Wechsler, 1997) or the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; J. N. Butcher, W. G. Dahlstrom, J. R. Graham, A. Tellegen, & B. Kaemmer, 1989) with Asian Americans. Issues surrounding the availability, reliability, and validity of assessment instruments must be addressed before extended discussions about the implication of test revisions for this population can take place. The authors review the current status of the WAIS-III and MMPI-2 with Asian Americans with respect to their availability, reliability, and validity, including reasons why Asian Americans have been severely underrepresented in validation studies. The authors argue for the need to collect data on the use of standardized assessment instruments with Asian Americans and conclude with recommendations for the inclusion of this population in future test revision projects.

Test revision is an enormously costly undertaking. Tests are revised for a number of reasons, but two of the foremost considerations are the reduction of bias and improved sampling of ethnic minorities. We have been asked to comment on major cross-cultural issues in test revisions pertinent to Asian American adults. However, this is a difficult task given that so little empirical information is available concerning the use of standardized assessment instruments with Asian Americans. In some ways, any consideration of test revision issues facing this population is superseded by more pressing concerns regarding the adequacy of existing instruments for use with this population. Why is test revision a "nonissue" for Asian Americans at this time?

We assert that before fruitful discussions about implication of revisions can take place three aspects of the assessment instruments in question must be established: their availability, reliability, and validity with a population group. On all three counts, there are no or inadequate data for Asian Americans for the majority of commonly used standardized assessment instruments used in the U.S. However, a consideration of test revision issues brings to focus the urgent need for good data on various tests with Asian Americans. We first discuss the current status of standardized tests with respect to Asian Americans and then discuss why it is important to improve the state of assessment practice with this population. Next, we argue that test revision efforts can be viewed as opportunities to improve testing with Asian Americans. In the final section, we make recommendations regarding future revisions of standardized assessment instruments.

Sumie Okazaki, Department of Psychology, University of Wisconsin—Madison; Stanley Sue, Department of Psychology, University of California, Davis.

Preparation of this article was supported in part by Grants K01 MH01506 and R01 MH44331 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to Sumie Okazaki, who is now at the Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent to okazaki@uiuc.edu.

We center our discussion around two of the most widely used instruments in the field, the revisions of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1967) and the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), although the issues we raise apply to other tests or to other groups as well. It must be acknowledged that both instruments have been criticized on a number of grounds. For example, Helmes and Reddon (1993) argued that, although the MMPI-2 is an improvement over the original MMPI, there are still major theoretical and structural concerns with both scales. Their criticisms include suspect criteria for diagnostic categories, heterogeneity of content within scales, and redundancy of items among the scales. However, given that each instrument has undergone a major revision within the past decade, together they serve as illustrative starting grounds for our discussion.

Current Status of Assessment With Asian Americans

As previously stated, the tests must be available, reliable, and valid for a given population before we can engage in meaningful discussions about the implications of test revisions for the population. In this section, we outline the current status of assessment with Asian Americans in these three regards.

Availability

Are major instruments such as the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and WAIS-III (Wechsler, 1997) available for use with all Asian Americans? Here, the English fluency of each Asian American testing client plays a large role in determining whether a given assessment instrument can be administered in a standard manner. According to the 1990 census, nearly two thirds of the Asian American population was born overseas (Ong & Hee, 1993), and thus many Asian Americans are not native English speakers. There is a wide range of language capabilities and preferences, with some Asian Americans being monolingual in the native Asian language, while many

are bilingual with varying degrees of facility in English and their native language. The availability of a translated version of an instrument is of primary concern with monolingual clients and bilingual clients with limited English skills.

With monolingual Asian American clients, the choices for the clinician are limited. Some clinicians have offered strategies for minimizing translation error inherent in on-the-spot translation of test items with monolingual Asian American clients (e.g., Yee, 1997). However, López and Romero (1988) view the practice of translating test items during a testing session as failing to meet the *Standards for Educational and Psychological Testing's* (1985) mandate to establish the reliability and validity of test translation before its use. It would be ideal if an Asian American client for whom English is not his or her best language could be tested by a trained bicultural, bilingual test administrator in the native language of the testee with translated versions of instruments that have established psychometric properties. The reality is far from this ideal, where an assessor often searches urgently for any translated version of a particular test in the specific Asian language or the dialect of the presenting client. Even if the assessor successfully determines that the test has been translated into the desired language, it may only be available overseas.

Various versions of the WAIS have been translated and standardized in only a handful of Asian nations. For example, a Japanese edition of the WAIS-R (Wechsler, 1981) was standardized in 1989 and published in 1990 (Shinagawa, Kobayashi, Fujita, & Maekawa, 1990). The mainland Chinese version of the original WAIS (Wechsler, 1955) was published as the Wechsler Adult Intelligence Scale-Revised for China (WAIS-RC; Gong, 1982). There appears to exist a Korean adaptation of the Wechsler-Bellevue scale (Koh, 1954). A Hindi version of the WAIS-R also appears to exist, as attested to by a number of publications in journals in India using this instrument with the Indian population (e.g., Pershad, Verma, & Randhawa, 1980; Srivastava & Shukla, 1988). However, it is uncertain whether these Asian language adaptations of the WAIS instruments are available in the U.S.

Moreover, even when a clinician is successful in obtaining and administering the specific Asian language version of the WAIS to a bilingual or monolingual Asian American client, interpretive difficulties follow. In discussing the practical aspects of testing monolingual Chinese American clients, Yee (1997) writes that older or original versions of an instrument are often used because they are the only versions available in Chinese translation. Such a practice creates a problem of comparing a given client's performance on a test against an outdated norm. There is evidence that IQ scores tend to rise over the years so that the IQ scores of an individual who is tested with an older version of the test (e.g., WAIS) would overestimate their true scores (see Flynn, 1999). Thus a monolingual Chinese American client's score on a Chinese translation of the WAIS administered in the 1990s may not be comparable to the WAIS norms that were established in the 1950s.

There are no studies that have directly examined the effect of the language used in the assessment on the outcome of clinical assessment with Asian Americans. However, clinical assessment studies with bilingual Spanish-speaking adults suggest that the language that is used in the assessment, as well as the ethnicity of the assessor, have nontrivial effects on the outcome of assessment. For example, Malgady and Constantino (1998) conducted psychiatric interviews with 148 Spanish-dominant bilingual first generation

Hispanic patients with schizophrenia, depression, and anxiety disorders. The patients matched on *DSM-IV* Axis I diagnosis, gender, place of birth, and Hispanic nationality were randomly assigned to one of four conditions: English interview by a non-Hispanic clinician, Spanish interview by a Hispanic clinician, English interview by a Hispanic clinician, or bilingual interview by a Hispanic clinician. Malgady and Constantino found that Hispanic clinicians tended to rate patients as more pathological than non-Hispanic clinicians, particularly when the interviews were conducted in Spanish or in both English and Spanish. For those Hispanic patients with schizophrenia, the magnitude of the severity rating differences on thought disorder symptoms between the bilingual interview condition and other conditions was greater than one standard deviation unit.

No direct studies of language match on assessment with Asian Americans have been conducted, but there is some indirect evidence that ethnic and language match between clinician and patients are associated with a number of related variables. For example, in a study of a community mental health system in Los Angeles, S. Sue, Fujino, Hu, Takeuchi, and Zane (1991) found that the combined match of ethnicity and language between clinician-patient pairs was consistently and significantly related to treatment length, dropout rate, and a global rating of functioning among Asian American clients whose primary language was not English. Taken together, the various studies involving language match suggest that it is entirely possible that some interactive effects of clinician-patient ethnicity and language match operate in testing situations with bilingual Asian Americans. Consequently, the fact that a clinician could test bilingual Asian Americans with a standard English version of a particular test would not preclude interpretive concerns.

In contrast to the lack of availability of WAIS-III or its predecessors in Asian languages, the MMPI and its successor, the MMPI-2, have been widely translated and adapted into various Asian languages. Their cross-cultural popularity is evident in the publications of compendium volumes on cross-cultural adaptation of these instruments (Butcher, 1996; Butcher & Pancheri, 1976). The MMPI-2 has been translated into Japanese (Shiota, Krauss, & Clark, 1996), Korean (Han, 1996), Chinese (Cheung, Song, & Zhang, 1996), Thai (Pongpanich, 1996), Vietnamese (Tran, 1996), and Hmong (Dienard, Butcher, Thao, Vang, & Hang, 1996). The audiocassette version of the Hmong translation of the MMPI-2 is also available so that it can be administered aurally to illiterate monolingual Hmong individuals (Dienard et al., 1996).

Of those Asian translations, the Vietnamese and the Hmong versions were developed primarily for use with Vietnamese and Hmong refugees residing in the U.S. (Dienard et al., 1996; Tran, 1996), while the others were adapted primarily for use in Asia. Butcher (1996; Butcher, Lim, & Nezami, 1998) credits the extensive development of various language versions of the MMPI-2 to a number of factors, including globalization of psychology and technological advances in communication. Particularly relevant to the present discussion are the changes made in the MMPI revision to facilitate cross-cultural applications. Specifically, the items in the MMPI-2 were revised or added with ease of translation in mind; the new norms were based on a more diverse sample of Americans that may have been a closer match to normative samples from other countries; and the revised version includes a broader range of scales to assess clinical problems that may be of

interest to the international mental health community. With respect to the availability of the various Asian language versions of the MMPI-2, an appendix in Butcher's (1996) compendium volume provides a list of the names and addresses of the translators and international researchers. Furthermore, because the Hmong version of the MMPI-2 is licensed by the University of Minnesota Press, it is available from a test distribution company (National Computer Systems, Inc.). For clinicians or researchers faced with monolingual Asian American clients or research participants, many Asian language translations of the MMPI-2 do exist, although some language versions may be more readily available in the U.S. than others.

Reliability

There are no published data on the reliability of WAIS or its subsequent revisions with Asian Americans. With the MMPI-2, the Vietnamese adaptation (Tran, 1996) and the Hmong adaptation (Deinard et al., 1996) have been conducted with Vietnamese Americans and Hmong Americans residing in Minnesota. Test-retest reliability figures are available for these two adaptations as a part of the bilingual retest study. A small sample of Vietnamese Americans ($n = 32$) and Hmong Americans ($n = 35$) who were bilingual in English and their respective languages were administered the English and the Vietnamese or Hmong versions of the MMPI-2 on two separate occasions. Tran (1996) reports the test-retest reliability for a 2–8 day interval ranged from .51 to .87, while Deinard et al. (1996) report the 7-day test-retest reliability as .38 to .80. These reliability figures may be underestimates given that they represent reliability across two different languages and across time. S. Sue, Keefe, Enomoto, Durvasula, and Chao (1996) collected the MMPI-2 data with English-speaking Asian American university students, and they reported comparable internal consistency estimates (alpha coefficients) for the validity and clinical scales for the two ethnic groups.

Validity

Again, there are no published studies of validity of Wechsler tests with Asian American adults, but there is one study of relevance to our discussion involving Asian American children. Saccuzzo and Johnson (1995) evaluated the use of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) and the Standard Raven Progressive Matrices (SPM; Raven, 1938) in the selection of children for giftedness. The performance of these two instruments for identifying gifted children was examined in the context of intervention and program evaluation involving school district-wide use of tests. Of the 16,985 children in the particular school district who were referred and evaluated for the gifted program, 483 were Asian, 104 were Pacific Islander, 1,419 were Filipino, and 958 were Indochinese. The researchers concluded that the WISC-R and the SPM were comparable in their abilities to predict academic achievement (as measured by standardized tests) and that there were no ethnic differences in the predictive validity of the tests. Interestingly, the researchers observed that the SPM (which can be administered to non-English speaking individuals due to its nonverbal stimuli) was effective in selecting gifted Indochinese children almost exactly in proportion to their numbers in the school district. Based on this one study of

primarily English-speaking Asian American children, the WISC-R was shown to have predictive validity for academic achievement.

There have been some efforts toward establishing the validity of the MMPI-2 with Southeast Asian refugee groups in the U.S. In the Vietnamese MMPI-2 translation and validation study, 193 Vietnamese American adults (94 males) in the St. Paul and Minneapolis areas were administered the Vietnamese version of the MMPI-2 (Tran, 1996). The sample consisted largely of unemployed adults (63%) who were attending college or vocational or technical training, over half of whom had lived in the U.S. for 2 years or less. When the mean profile of the Vietnamese male and female samples were plotted against the American norms, both profiles were within the normal range, although scales F and 8 were slightly higher than the American norms. Factor analysis of the Vietnamese MMPI-2 revealed a factor structure similar to that of the American normative sample. Tran concluded that an average Vietnamese American male or female appeared similar to the average American male or female on the MMPI-2, with the exception of cultural factors such as modesty and attitudes toward sex and mental health treatment. According to Butcher, Lim, and Nezami (1998), standardization and discriminant validity studies for the Hmong translation of the MMPI-2 are in progress but no validity data are available for the Hmong Americans as of this date. In a study using the English MMPI-2 with 133 Asian American and 91 White American university students, S. Sue et al. (1996) plotted the mean profiles of both groups against the American norm and found that highly acculturated Asian Americans and White Americans had very similar profiles. Of interest, both Tran (1996) and S. Sue et al. (1996) reported that the greatest differences from the American normative profile were seen in the Vietnamese American and Asian American individuals who immigrated most recently and were least acculturated to the United States.

Another issue raised in the discussion of the validity of the standardized assessment instruments is the nature of the standardization sample. Standardization samples for the recent revisions of the Wechsler tests and the MMPI have used the proportionate sampling strategy, which involves conducting stratified random sampling using ethnic groups and other demographic variables as strata. This sampling approach was employed to ensure that the standardization samples included representative proportions of the children or adults, stratified by age, gender, race/ethnicity, education level (for adults), parent education (for children), and geographic regions. The WAIS-III stratification was based on the U.S. census data from 1995 (Psychological Corporation, 1997), while the MMPI-2 stratification was based on the U.S. census data from 1980 (Butcher et al., 1989). For Asian Americans, such a sampling procedure results in extremely small numbers of Asian Americans being included in the standardization sample. The WAIS-III standardization sample for each age group was designed to reflect the proportion of Whites, African Americans, Hispanics, and "other" racial/ethnic groups, and no exact figures for the percentage or raw sample size of Asian Americans are cited in the technical manual (Psychological Corporation, 1997). In the MMPI-2 standardization sample, only 6 Asian American males and 13 Asian American females (constituting less than 1% of the total sample) were included (Butcher et al., 1989), while Asian Americans represented less than 2% of the U.S. population in 1980 and 3% of the U.S. population in 1990 (Kitano & Daniels, 1995).

What is the implication of the proportionate sampling for the use of these instruments with Asian Americans? Some have argued, based on observational or animal genetic models, that this population-proportionate sampling practice in combination with the classical item selection procedures for test construction may favor the groups that have larger representation in the construction sample (e.g., Harrington, 1975, 1988; Jensen, 1980). Fan, Willson, and Kapes (1996) empirically examined the viability of the proportionate-sampling test bias claim by experimentally modeling the differential representation versus maximum representation of four major ethnic groups in test construction samples, selecting test items based on item-total correlation for each test construction sample and comparing ethnic group performances based on the various resulting versions of the test. The researchers concluded that there were no empirical bases for bias against ethnic groups with smaller representation in the test construction sample, at least for achievement-type tests (e.g., math and reading tests). However, similar empirical tests of proportionate sampling for intelligence tests or objective personality tests have not been carried out.

The discussion of the nature of the standardization samples is relevant to the validity of applying the various norms for the WAIS-III or the MMPI-2 to individual Asian American cases. That is, even if an assessment instrument was reliably administered to a given Asian American client in the appropriate Asian language, the interpretation of his or her scores relative to the normative population remains an issue. Asian Americans are also likely to be different from their overseas counterparts, and using Asian norms from Asia may not be appropriate. As noted by S. Sue, D. W. Sue, L. Sue, and Takeuchi (1995), immigrants to the United States are not representative of those who remain in their native countries. Some immigrants come from educated, upper-class backgrounds while others have led impoverished lives in their homelands and flee to the United States for better opportunities. Immigrants and refugees encounter unique experiences in this country, such as exposure to different cultural values, English proficiency problems, minority status (including racial or ethnic stereotypes), prejudice, discrimination, and a reduction in available social supports. The immigrant population in this country may also exhibit greater demographic and cultural diversity than those overseas. For example, the Chinese American population is far more diverse than the Chinese population clusters in Asia that have been studied (e.g., Hong Kong Chinese, Chinese in Beijing, etc.). Chinese in America are composed of both native and foreign-born individuals whose families migrated directly to the U.S. from Asian nations such as mainland China, Taiwan, Hong Kong, Singapore, and Vietnam as well as secondarily from South America, Latin America, and other nations. Chinese Americans speak many dialects and languages, are exposed to American values, and are members of a minority group. All of these produce a very heterogeneous Chinese American population—more so than, say, Chinese in mainland China, who are more ethnically, linguistically, and culturally homogenous.

The same is true of nearly all Asian American groups, where greater heterogeneity exists for the group in the U.S. than in their homeland. This is an important point to underscore, as it means that measures (such as the MMPI-2) that have been validated with Chinese in specific sites in China may not have good validity for Chinese residing in the U.S. Furthermore, if great diversity exists even among Chinese Americans alone, then combining all Asian

ethnic groups would result in an enormously heterogeneous population group. Without empirically and systematically examining the validity of both U.S. and overseas Asia norms to Asian Americans, we do not know if Asian Americans (or subsamples of this population) need to be normed as a separate standardization sample.

Lack of Data

While there are a number of studies of the reliability and validity of Wechsler scales or the MMPI tests with African American and Hispanic populations (e.g., Fantoni & Rogers, 1997; Hall, Bansal, & Lopez, 1999; Timbrook & Graham, 1994), there are virtually no data on psychometric properties of standardized instruments with Asian Americans. In fact, Asian Americans continued to be classified as "Other" along with Native Americans and multiracial individuals in the WAIS-III standardization sample (Psychological Corporation, 1997). Even the category of "Asian" masks tremendous heterogeneity among the various Asian groups (e.g., Chinese, Filipino, Japanese, Korean, Vietnamese, Indian, etc.).

Why is there a failure to include and distinguish among the different Asian American groups in test standardization efforts? We believe that the primary reason is the sheer sampling and methodological difficulty inherent in collecting assessment data with Asian Americans (Okazaki & S. Sue, 1995). For example, in a study that necessitated obtaining a representative sample of Chinese Americans in Los Angeles County for an epidemiologic study, Takeuchi et al. (1998) sampled nearly 20,000 households. This study required a great deal of funding and effort (e.g., translating and pretesting the instruments, training bilingual interviewers, and visiting a large number of households) in order to locate a representative sample of 1,700 Chinese Americans. Given that most researchers lack access to the resources necessary to conduct such large-scale studies, Asian Americans are notably missing from assessment studies with normative population and clinical populations. Even in large-scale standardization studies for the widely used instruments, Asian Americans are typically not oversampled. For example, the WAIS-III technical manual (Psychological Corporation, 1997) states that the WAIS-III was administered without the usual discontinue rules to an additional group of 200 African American and Hispanic individuals so that item bias analyses for these two ethnic groups could be conducted. However, similar data on additional Asian Americans and Native Americans were not collected. It is likely that Asian Americans were not oversampled in the recent major test revision efforts due to the difficulties of locating Asian American who are small in number and also geographically scattered and because of concerns about the level of English fluency of some Asian Americans. Clearly, excluding Asian Americans at the "ground floor" of test revision projects (e.g., by failing to oversample this population for reliability and validity studies) contributes to the lack of psychometric and cross-cultural equivalence data for the standardized assessment instruments.

In addition to the notable absence of Asian Americans from large-scale standardization samples, there are virtually no data using standardized assessment instruments with clinical samples of Asian Americans. Here, a different set of forces may be playing a role. We speculate that due to lack of familiarity with psychological testing practices, fewer Asian Americans present themselves

or are referred to testing than other ethnic groups.¹ In regard to cognitive abilities testing, Asian Americans may be less likely to seek learning disabilities testing on their own because of the stigma attached to mental deficiency, cultural values that emphasize effort over abilities, and community emphasis on achievement (Uba, 1994). As an anecdote, school psychologists in school districts with high Asian American enrollment have approached one of the authors to convey their frustration with Asian American parents of English-speaking children referred for psychological testing. According to these psychologists, the Asian American children in question were identified as having academic difficulties for which testing may have been able to suggest appropriate interventions, yet the parents refused to consent to having their children tested, presumably due to their fear of a stigma associated with possible outcomes such as low IQ or learning disability. Thus, cultural barriers to testing may exist for Asian Americans who may benefit from such services.

Further, Asian Americans are less likely to be represented in institutional settings such as forensics or the veterans' services where psychological assessment research is conducted. For example, there are studies using the MMPI-2 or the Wechsler scale with African American male veterans (Frueh, Gold, de-Arellano, & Brady, 1997), with African American and Hispanic male offenders (Hiltonsmith, Hayman, & Kleinman, 1984), and African American men court-ordered for forensic psychological evaluations (Ben-Porath, Shondrick, & Stafford, 1995), but no comparable samples of Asian Americans have been studied with respect to assessment instruments. Finally, even when Asian Americans do submit to clinical assessment, individual clinicians may lack either cultural knowledge or access to resources for conducting culturally competent assessment that includes standardized instruments. Alternatively, those clinicians who regularly conduct psychological testing with Asian American clients may not engage in the type of data collection that results in empirical research publication. For a number of reasons, then, a remedy for lack of research using standardized assessment instruments with Asian Americans would require considerable research expertise, funding, and organizational structure.

Why Asian Americans Must Be Included

On the one hand, it is understandable why Asian Americans continue to be underresearched; methodological challenges are formidable. On the other hand, there are many compelling reasons why it is imperative for the assessment field to start collecting data on standardized assessment instruments with Asian Americans. In this section, we will outline three major reasons why we need to establish test reliability and validity with Asian Americans: (a) The growing size of the Asian American population will be reflected in an increasing number of Asian Americans who need assessment services, (b) the choice between continuing to use standardized assessment with no reliability or validity data with Asian Americans and not using any assessment instruments at all constitutes an ethical dilemma, and (c) the scientific importance of establishing the generalizability of theories underlying assessment instruments across culture and ethnicity.

Changing Demographics, Changing Needs

The population growth of Asian Americans is expected to continue well into the next century. By the year 2020, Asian Americans are expected to show a population increase of 145% to 177% from their number in 1990 (Ong & Hee, 1993). Asian Americans have already become a sizable presence in states like California and Texas. Given that the majority of Asian Americans are immigrants and given that Asian Americans are not likely to become wholly assimilated into the American population in the foreseeable future (if ever), the present lack of cultural validity of assessment instruments with Asian Americans is expected to worsen if no actions are taken.

Ethics in Testing

The American Psychological Association's *Standards for Educational and Psychological Testing* (1985) and the ethical standards of the American Association for Counseling and Development (Corey, 1988) are clear in stating the necessity of establishing the validity of tests and assessments in the evaluation of individuals who come from different cultures. The reason for this is also clear: Culture and linguistic factors often affect the entire interpretation of results by those who conduct assessment (Sandoval, 1998). Unfortunately, the almost complete lack of normative information on Asian Americans (Gray-Little & Kaplan, 1998) has only resulted in recommendations to be cautious in the interpretation of test results with Asian Americans. Rather than merely continuing with this admonishment, the assessment community must begin to conduct reliability and validity studies of standardized tests with Asian Americans and other cultural linguistic groups so that ethical standards may be realized in more assessment cases with members of these population groups.

Scientific Importance

In arguing for empirical studies of Asian Americans with standardized assessment tools, we are not stating that ethnic or cultural differences on psychological constructs should be valued over ethnic or cultural similarities. Indeed, Butcher (1996) argues that the bulk of international studies of the MMPI and the MMPI-2 in various nations have pointed to essential similarities of abnormal

¹ Underutilization of psychological services among Asian Americans is not limited to assessment. In general, Asian Americans have been found to underutilize public mental health services (S. Sue, Fujino, Hu, Takeuchi, & Zane, 1991). It is possible that Asian Americans are underrepresented in mental health services because they are a relatively healthy subpopulation in the U.S. However, recent data from a psychiatric epidemiological study of Chinese Americans suggest a complex picture of the rates of mental disorders; the 12-month prevalence rate of *DSM-III-R* major depression was lower than the rate in the comparable epidemiological surveys in the U.S. (Takeuchi et al., 1998), yet there was a significant number of Chinese Americans who met diagnostic criteria for neurasthenia (a syndrome characterized by fatigue or weakness accompanied by a variety of physical and psychological symptoms, still widely accepted in modern-day Chinese societies) (Zheng et al., 1997). There are also a number of cultural and linguistic barriers to seeking mental health services in the U.S. for Asian Americans (Uba, 1994) that may contribute to relative underrepresentation of Asian Americans in mental health services.

behavior across diverse cultures. However, the implicit assumption of ethnic similarity and the applicability of standardized tests or the U.S. norms are questionable (Reid, 1994), as those assumptions have yet to be empirically examined. In the past decade, psychology as a field has begun to confront the increased multiculturalism and globalization of research and the accompanying realization that the model of psychological processes (e.g., the self) held in the U.S. is "quite simply not the one that is held by the majority of the people in the world" (Kitayama & Markus, 1995, p. 367). Malgady (1996) discussed this issue in terms of the prevailing null hypothesis (i.e., the assumption of cross-cultural uniformity in assessment and diagnosis) in the absence of a definitive empirical basis for showing that there are cultural biases in mental health assessment and psychiatric diagnosis of ethnic minorities. Malgady argued that the traditional null hypothesis must be challenged, because the consequences associated with its incorrect retention (Type II error leading to disservice to minority clients) may be greater than its incorrect rejection (Type I error leading to the mental health system potentially misallocating resources).

For the science of psychology to truly reflect all aspects of its tenets (i.e., to understand human psychological functioning, not just White middle-class Americans' psychological functioning), the inclusion of Asian Americans in testing and other psychological research is imperative. S. Sue (1999) has also indicated that research on ethnic and culturally different populations is important not only to understand these populations but also to advance scientific knowledge. It is only through the study of diverse populations that the generality (or lack of generality) of psychological theories and principles can be established. Furthermore, diverse populations are likely to increase the range of variables being studied. For example, it has been through the studies of ethnic and linguistic minority groups in this country (e.g., Malgady & Constantino, 1998) that the field has begun to understand the importance of diagnostic and assessment bias.

Recommendations for Test Revisions

Test Revisions as the Best Chance for Collecting Large Normative Data

The recent MMPI and WAIS revisions have made increasing efforts to eliminate bias in testing with ethnic minorities and, to this end, oversampled African American and Hispanic individuals in their standardization sampling. The availability of a relatively large sample of African Americans through the MMPI-2 restandardization project has made some validity studies possible (e.g., Timbrook & Graham, 1994). As evidenced by the collapsing of Asian Americans with Native Americans, Alaskan Natives, Pacific Islanders, and presumably multiracial individuals into the "Other" category in the WAIS-III standardization sample (Psychological Corporation, 1997), Asian Americans as an ethnic group have been largely ignored by the assessment community. We strongly urge the test developers to include Asian Americans as a part of similar future revision projects.

Test Revisions as Opportunities to Test Moderator Variables

Multicultural assessment expert Richard Dana (1993, 1998) has long advocated the use of moderator variables such as accultura-

tion² level in interpreting the results of standardized assessment tests with cultural minority groups. Indeed, the data on the MMPI-2 with small samples of Vietnamese American community residents (Tran, 1996) and Asian American university students (S. Sue et al., 1996) indicated that length of residency and level of acculturation to the American culture appear related to the differential elevations on the MMPI-2 scales when plotted against the U.S. norm. In both studies, the least acculturated groups of Asian Americans were more likely to score in the more pathological directions on various scales. However, such data in and of themselves were not able to clarify whether the scale elevations among the recently immigrated Asian Americans reflect greater distress due to acculturative stress or cultural response sets (S. Sue et al., 1996). A greater understanding of the relationship between acculturation and responses on standardized assessment instruments is needed. What adds to the complexity of the problem at hand are the facts that acculturation is a complex, multidimensional, and time-varying construct (Magaña et al., 1996) and that no empirical data are available regarding the exact nature of the relationship between various aspects of acculturation and standardized assessment instruments. Pilot studies that lead to test revisions may provide a perfect setting for studying how specific indices of acculturation may be associated with the constructs being assessed by the standardized instruments.

Test Revisions as Collaborative Opportunities

In her 1992 paper, Helms decried the poor state of knowledge with regard to the cultural equivalence of standardized test of cognitive abilities. She argued that the assessment field has not done a very good job in conceptualizing and examining the cultural equivalence issue in existing intelligence tests for ethnic minorities. Furthermore, Helms argued that the very small number of ethnic minority psychologists and psychometricians should not have to assume the burden for the entire field. The assessment community faces some difficult but necessary tasks that must be accomplished in order to deliver competent assessment practices with ethnic minorities. We strongly concur with Helms's challenge: "Who, then, will 'well develop' these culturally sensitive measures of cognitive abilities if not those renowned psychologists and psychometricians whose careers have flourished under the Eurocentric tradition in testing?" (p. 1099). At the same time, we recognize that research with ethnic minorities poses specific methodological challenges (Okazaki & S. Sue, 1995), and that psychologists, psychometricians, and test publishers may not always have the necessary expertise or access to the targeted ethnic minority population. However, as demonstrated by the recent success of the Chinese American Psychiatric Epidemiological Study project (CAPES; Takeuchi et al., 1998), reliable and valid large-sample data on psychological variables can be collected with Asian American communities using sophisticated multi-stage probability sampling methodologies.

² Acculturation is a process through which the attitudes and/or behaviors of person from one culture are modified as a result of contact with a different culture (Moyerman & Forman, 1992). In the case of Asian Americans, the concept of acculturation refers to the extent to which they acquire attitudes and behavior patterns consistent with those of the mainstream American culture.

We raise the concern that if the test developers or the test publishers do not take the lead or are not heavily involved in translating and conducting cross-cultural validity studies with various Asian American populations, the burden falls on individual researchers who may not have the funding or the resources to work in a timely manner and to follow the recommended rigorous and costly test adaptation procedures such as those outlined in Geisinger (1994). Often, by the time individual researchers have conducted cross-cultural adaptation studies with a given assessment instrument, the revised version of the same test may well be under development. Thus, it would be in everyone's best interest if scientists with specific expertise in sampling and testing Asian Americans are included in future test revision projects. Further, for truly collaborative relationships to develop between the mainstream test developers and ethnic researchers, the cultural expert must be brought in at the design stage of test revision projects rather than as occasional consultants. The recent publication of the Hmong adaptation of the MMPI-2 (Dienard et al., 1996) by the University of Minnesota Press is an excellent example of a collaborative effort between one of the authors of the MMPI-2 (Butcher) and cultural experts.

Suggested Studies

The sheer methodological and conceptual challenges of collecting data to represent even the basic elements of the diversity of Asian Americans can be overwhelming. How can the test revision dilemma for Asian Americans and other heterogeneous populations be solved? We offer here examples of future research studies. In approaching the problem of the suitability of U.S. norms for Asian Americans, the research solution is to gather some normative data for Asian Americans. It would not be practical at first to gather a sizable sample of all major ethnic Asian American ethnic groups (i.e., Chinese, Japanese, Korean, Filipino, etc.) representing the geographic, socioeconomic, and generational diversity of most Asian Americans in this country. However, some data with Asian Americans must be collected each time a standardized test is revised so as to not perpetuate the outdated norms problem for this population. As an initial step, it may be feasible to select and oversample two Asian ethnic groups so as to minimize the costs of translations and data collection and to guide future research efforts as to the utility of collecting ethnic-specific norms. The selection of the two population samples should be guided by both theoretical and practical concerns, and we offer two possible selection criteria below.

First, at least one of the samples should represent ethnic and linguistic groups that most frequently present for testing; for example, some school districts in large metropolitan areas may have a relatively large enrollment of Korean Americans or Chinese Americans or Vietnamese Americans, and children of these ethnic groups may present for testing with enough frequency. Consequently, normative data on children (or adults) from a similar ethnic background would be immediately useful. Data collection with school-age children may be more practical than adults, given that immigrant Asian American children are likely to have more English fluency than immigrant Asian American adults.

Second, the samples should be selected to maximize the inferences one can draw from the data. For example, to examine which aspect of acculturation (e.g., English language fluency, number of

years spent in U.S. schools, age at immigration, cultural values and attitudes, and behavioral patterns) among Asian Americans may be associated with variations in test scores or profiles, at least one of the samples should have distributions on the key acculturation indices. Another example of a hypothesis that can be tested is interethnic differences—are the tests equally reliable or equally valid in various Asian ethnic groups (e.g., Chinese Americans vs. Japanese Americans), and what factors significantly predict Asian ethnic group differences? To answer such a question, a researcher may select a Chinese American sample from suburban areas consisting of middle-class individuals to be contrasted with a Hmong refugee sample from rural areas. Or two Chinese American samples may be selected, one from a middle-class suburban setting where English is the predominant language and one from an urban center such as a Chinatown in a metropolitan city where English is not the predominant language. A Chinese American ethnic group was selected as the focus of the first psychiatric epidemiological study with Asian Americans (Takeuchi et al., 1998; Zheng et al., 1997) because this ethnic group has a more even representation of U.S.-born and foreign-born individuals compared to other Asian ethnic groups. Similar selection criteria based on the knowledge of various Asian ethnic communities' demographic and cultural characteristics can be applied to assessment research.

Additionally, to maximize the inferences that can be drawn based on limited data, we recommend collecting culturally relevant data concerning the testing conditions (e.g., ethnicity of the clinician, language used in testing, and how closely the standardized test protocol was followed), the demographic and cultural characteristics of each individual (e.g., socioeconomic status, gender, age, and acculturation measure), and the demographic and cultural characteristics of the community from which the individuals are drawn (e.g., ethnic composition and availability of English as a Second Language classes). Recruitment and data collection may capitalize on natural population clusters of the targeted Asian American individuals (e.g., school districts with high Asian American enrollment and outpatient clinics catering to Asian American clients). Once some normative data on standardized assessment instruments are available for some Asian American ethnic groups, assessment research and practice with Asian Americans will make significant progress. For example, if the ethnic-specific norms for the two Asian American ethnic groups are not different from the U.S. norm, one may tentatively conclude that Asian American individuals similar to the two targeted groups (and perhaps other Asian ethnic groups) may be referenced against the standard norm. In contrast, if the Asian ethnic-specific norms are substantially different from the standardization norm, further research would explore conceptual or methodological sources of the group differences that may suggest remedies for future test revision efforts.

In conclusion, there is a serious lack of research on the reliability and validity of assessment instruments for Asian American populations. The paucity of research is attributable to several factors that pose practical and theoretical challenges to conducting research with these populations, such as the relatively small numbers, the heterogeneous ethnic groups that comprise the population, and the different languages spoken by Asian Americans. Nevertheless, validation of test instruments is crucial, given that Asian Americans are likely to be increasingly represented in school, mental health, or other settings where standardized assessments are conducted. Test revisions provide a prime opportunity to study

culturally diverse populations. Small-scale studies such as those suggested here would yield rich data that can help to test the generality of psychological theories and constructs being measured by the assessment instruments undergoing the revision. Thus, on practical, ethical, and scientific grounds, we recommend the inclusion of Asian Americans in future test revision efforts.

References

- Ben-Porath, Y. S., Shondrick, D. D., & Stafford, K. P. (1995). MMPI-2 and race in a forensic diagnostic sample. *Criminal Justice and Behavior*, 22, 19-32.
- Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring. Minnesota Multiphasic Personality Inventory-2: MMPI-2*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Lim, J., & Nezami, E. (1998). Objective study of abnormal personality in cross-cultural settings: The Minnesota Multiphasic Personality Inventory (MMPI-2). *Journal of Cross-Cultural Psychology*, 29, 189-211.
- Butcher, J. N., & Pancheri, P. (1976). *Handbook of cross-national MMPI research*. Minneapolis: University of Minnesota Press.
- Cheung, F. M., Song, W., & Zhang, J. (1996). The Chinese MMPI-2: Research and application in Hong Kong and the People's Republic of China. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 137-161). Minneapolis: University of Minnesota Press.
- Corey, G. (1988). *Issues and ethics in the helping professions*. Pacific Grove, CA: Brooks Cole.
- Dana, R. H. (1993). *Multicultural assessment perspectives for professional psychology*. Boston: Allyn & Bacon.
- Dana, R. H. (1998). *Understanding cultural identity in intervention and assessment*. Thousand Oaks, CA: Sage.
- Dienard, A., Butcher, J. N., Thao, U. D., Vang, S. H. M., & Hang, K. (1996). Development of a Hmong translation of the MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 194-205). Minneapolis: University of Minnesota Press.
- Fan, X., Willson, V. L., & Kapes, J. T. (1996). Ethnic group representation in test construction samples and test bias: The standardization fallacy revisited. *Educational and Psychological Measurement*, 56, 365-381.
- Fantoni, S., & Rogers, R. (1997). Spanish versions of the MMPI-2 and PAI: An investigation of concurrent validity with Hispanic patients. *Assessment*, 4(1), 29-39.
- Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, 54, 5-20.
- Frueh, B. C., Gold, P. B., de-Arellano, M. A., & Brady, K. L. (1997). A racial comparison of combat veterans evaluated for PTSD. *Journal of Personality Assessment*, 68, 692-702.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6, 304-312.
- Gong, Y. X. (1982). *Manual for the Wechsler Adult Intelligence Scale: Chinese Revision*. Changsha, Hunan, China: Hunan Medical College.
- Gray-Little, B., & Kaplan, D. A. (1998). Interpretation of psychological tests in clinical and forensic evaluations. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 141-178). Washington, DC: American Psychological Association.
- Hall, G. C., Bansal, A., & Lopez, I. R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment*, 11, 186-197.
- Han, K. (1996). The Korean MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 88-136). Minneapolis: University of Minnesota Press.
- Harrington, G. M. (1975). Intelligence tests may favor the majority groups in a population. *Nature*, 258, 708-709.
- Harrington, G. M. (1988). Two forms of minority-group test bias as psychometric artifacts with an animal model (*Rattus norvegicus*). *Journal of Comparative Psychology*, 102, 40-407.
- Hathaway, S. R., & McKinley, J. C. (1967). *Minnesota Multiphasic Personality Inventory manual* (Rev. ed.). New York: Psychological Corporation.
- Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI-2. *Psychological Bulletin*, 113, 453-471.
- Helms, J. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- Hiltonsmith, R. W., Hayman, P. M., & Kleinman, P. (1984). Predicting WAIS-R scores from the Revised Beta for low functioning minority group offenders. *Journal of Clinical Psychology*, 40, 1063-1066.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Kitano, H. H. L., & Daniels, R. (1995). *Asian Americans: Emerging minorities* (2nd ed). Englewood Cliffs, NJ: Prentice Hall.
- Kitayama, S., & Markus, H. R. (1995). Culture and self: Implications for internationalizing psychology. In N. R. Goldberger & J. B. Veroff (Eds.), *The culture and psychology reader* (pp. 366-383). New York: New York University Press.
- Koh, S. (1954). A statistical evaluation of the Korean conversion of the Wechsler-Bellevue Intelligence Scale. *Studies in Psychology, Ewha Woman's University*, 1, 26-53.
- López, S. R., & Romero, A. (1988). Assessing the intellectual functioning of Spanish-speaking adults: Comparison of the EIWA and the WAIS. *Professional Psychology*, 20, 369-379.
- Magaña, J. R., de la Rocha, O., Amsel, J., Magaña, H. A., Fernandez, M. I., & Rulnick, S. (1996). Revisiting the dimensions of acculturation: Cultural theory and psychometric practice. *Hispanic Journal of Behavioral Science*, 18(4), 444-468.
- Malgady, R. G. (1996). The question of cultural bias in assessment and diagnosis of ethnic minority clients: Let's reject the null hypothesis. *Professional Psychology: Research and Practice*, 27, 73-77.
- Malgady, R. G., & Constantino, G. (1998). Symptom severity in bilingual Hispanics as a function of clinician ethnicity and language of interview. *Psychological Assessment*, 10, 120-127.
- Moyerman, D. R., & Forman, B. D. (1992). Acculturation and adjustment: A meta-analytic study. *Hispanic Journal of Behavioral Sciences*, 14, 163-200.
- Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, 7, 367-375.
- Ong, P., & Hee, S. J. (1993). The growth of the Asian Pacific American population: Twenty million in 2020. In *The State of Asian Pacific America: Policy issues to the year 2020* (pp. 11-23). Los Angeles, CA: LEAP Asian Pacific American Public Policy Institute and UCLA Asian American Studies Center.
- Pershad, D., Verma, S. K., & Randhawa, A. (1980). Adaptation of the verbal scale of the WAIS-R. *Psychological Research Journal*, 4(1), 41-48.
- Pongpanich, L. (1996). Use of the MMPI-2 in Thailand. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 162-174). Minneapolis: University of Minnesota Press.
- Psychological Corporation. (1997). *WAIS-III-WMS-III Technical Manual*. San Antonio, TX: Author.
- Raven, J. C. (1938). *Progressive matrices*. London: Lewis.
- Reid, P. T. (1994). The real problem in the study of culture. *American Psychologist*, 49, 524-525.
- Saccuzzo, D. P., & Johnson, N. E. (1995). Traditional psychometric tests

- and proportionate representation: An intervention and program evaluation study. *Psychological Assessment*, 7, 183-194.
- Sandoval, J. (1998). Testing in a changing world: An introduction. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 3-16). Washington, DC: American Psychological Association.
- Shinagawa, F., Kobayashi, S., Fujita, K., & Maekawa, H. (1990). *Japanese Wechsler Adult Intelligence Scale-Revised*. Tokyo: Nihon Bunka Kagakusha.
- Shiota, N. K., Krauss, S. S., & Clark, L. A. (1996). Adaptation and validation of the Japanese MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 67-87). Minneapolis: University of Minnesota Press.
- Srivastava, A., & Shukla, T. R. (1988). WAIS-R (Hindi, verbal) and WAPIS (Form-PR) in differential diagnosis of schizophrenia and organicity. *Indian Journal of Clinical Psychology*, 15(1), 42-44.
- Standards for Educational and Psychological Testing*. (1985). Washington, DC: American Psychological Association.
- Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, 54, 1070-1077.
- Sue, S., Fujino, D. C., Hu, L., Takeuchi, D. T., & Zane, N. W. S. (1991). Community mental health services for ethnic minority groups: A test of the cultural responsiveness hypothesis. *Journal of Consulting & Clinical Psychology*, 59, 533-540.
- Sue, S., Keefe, K., Enomoto, K., Durvasula, R. S., & Chao, R. (1996). Asian American and White college students' performance on the MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 206-218). Minneapolis: University of Minnesota Press.
- Sue, S., Sue, D. W., Sue, L., & Takeuchi, D. (1995). Psychopathology among Asian Americans. *Cultural Diversity and Mental Health*, 1, 39-54.
- Takeuchi, D. T., Chung, R. C., Lin, K., Shen, H., Kurasaki, K., Chun, C., & Sue, S. (1998). Lifetime and twelve-month prevalence rates of major depressive episodes and dysthymia among Chinese Americans in Los Angeles. *American Journal of Psychiatry*, 155(10), 1407-1414.
- Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2? *Psychological Assessment*, 6, 212-217.
- Tran, B. N. (1996). Vietnamese translation and adaptation of the MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp. 175-193). Minneapolis: University of Minnesota Press.
- Uba, L. (1994). *Asian Americans: Personality patterns, identity, and mental health*. New York: Guilford Press.
- Wechsler, D. (1955). *WAIS manual: Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.
- Yee, J. (1997). Psychological testing of monolingual Asian American clients. In E. Lee (Ed.), *Working with Asian Americans: A guide for clinicians* (pp. 364-387). New York: Guilford.
- Zheng, Y., Lin, K., Takeuchi, D., Kurasaki, K. S., Wang, Y., & Cheung, F. (1997). An epidemiological study of neurasthenia in Chinese-Americans in Los Angeles. *Comprehensive Psychiatry*, 38(5), 249-259.

Received April 29, 1999

Revision received September 8, 1999

Accepted October 4, 1999 ■