Cognitive–Intellectual Functioning of Spanish-Speaking Impaired and Nonimpaired Elderly: Implications for Culturally Sensitive Assessment

Steven Regeser López and I. Maribel Taussig
University of Southern California

On the basis of prior research (López & Romero, 1988), we examined whether using the Wechsler Adult Intelligence Scale-Revised (WAIS–R) can lead to an underestimation of Spanish-speaking older adults' functioning and whether using the Escala de Inteligencia Wechsler para Adultos (EIWA) can lead to an overestimation of this group's functioning. Spanish-speaking patients who met criteria for Alzheimer's Disease (AD) and Spanish-speaking nonimpaired elderly were tested with selected subtests from the WAIS–R and EIWA, and the Mini-Mental State Examination. The AD patients' activities of daily living were also assessed. For purposes of comparison, English-speaking nonimpaired older adults and AD patients were administered the same tests except that the Wechsler subtests were all taken from the WAIS–R. The results point out clear differences between impaired and nonimpaired groups, and differences and similarities between the two language groups. We interpret the differences between the language groups as reflecting test bias when using either the WAIS–R or EIWA in assessing Spanish-speaking elderly. The findings have implications for defining culturally sensitive psychological assessment.

The psychologist who chooses to assess Spanish-speaking individuals is confronted with a major challenge: to conduct psychological assessment in a culturally sensitive fashion. López et al. (1989) defined cultural sensitivity as carefully balancing general norms with culture-specific norms in judging a specific patient or client. Consider the example of an elderly woman with limited knowledge of English who performs poorly on an English language test such as the Wechsler Adult Intelligence Scale—Revised (WAIS–R; Wechsler, 1981). In adopting an etic perspective, a perspective that applies what are thought to be general norms, one may then interpret the test results as reflecting impairment. Compared with the English language standardization sample, this predominantly Spanish-speaking elderly woman is functioning at the same level as English-speaking elderly individuals who are impaired in their cognitive–intellectual functioning. In this case, however, it is possible that the below-average performance may reflect limited English language skills, an unfamiliarity with the American culture, a limited educational background, or all of these factors. Therefore, when applying etic norms, in this case WAIS–R norms, it seems possible that psychologists could underestimate the cognitive–intellectual functioning of Spanish-speaking adults.

To adopt an emic perspective, a culture-specific perspective, a psychologist might use the Spanish-language Wechsler test, Escala de Inteligencia Wechsler para Adultos (EIWA; Wechsler, 1968). The instrument was standardized on a large sample of Spanish-speaking adults from Puerto Rico. Given the language of the instrument and its standardization sample, it would appear to contain emic norms, norms most appropriate for Spanish-speaking adults in the United States. A close examination of the test suggests that it too may be limited in assessing Spanish-speaking adults.

López and Romero (1988) examined the comparability of the EIWA and its English predecessor, the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). They found that the two tests differ in many important ways, particularly in the conversion of raw scores to scale scores. The performance on a given subtest can result in a scale score difference of up to 5 points, depending on whether one applies WAIS or EIWA norms. For instance, a raw score of 16 on the Object Assembly, a subtest that is identical on the WAIS and the EIWA, translates into a scale score of 5 for the WAIS and 10 for the EIWA.

For some Spanish-speaking adults, the EIWA may be most appropriate, particularly for those with low educational levels, low occupational status, and rural backgrounds—the combined background of most of the EIWA standardization sample. The elderly Spanish-speaking woman who tested as "impaired" using English-language norms, for example, might likely be judged to be functioning within the average range. It is important to note, however, that the test's standardization took place over 23 years ago. Since then, the educational level of residents of Puerto Rico has improved. For instance, the percentage of Puerto Ricans 25 years and older who have completed at least a high school education has increased from 27% in 1970 to 39.6% in 1980 (U.S. Department of Commerce, 1984). Thus, it is likely that what was considered an average level of cognitive–intellectual functioning in 1965 reflects a lower level of functioning now. As a result, psychologists who...
Currently, test Spanish-speaking persons using the EIWA may judge persons with some impairment to be functioning within the average "normal" range. Further, persons with higher educational and occupational backgrounds than the standardization sample are likely to test rather high, not because they are necessarily high functioning, but because their backgrounds deviate from that of the standardization sample. In all, it seems possible that applying the apparent culture-specific norms of the EIWA, psychologists may be at risk to overestimate the intellectual functioning of some Spanish-speaking adults.

The main purpose of the present study is to directly test whether the cognitive-intellectual functioning of Spanish-speaking adults is prone to be underestimated when using the WAIS-R and overestimated when using the EIWA, as suggested by the present analysis. To do so, we examined data from a research project that is developing a Spanish-language neuropsychological test battery for the diagnosis of Alzheimer's Disease (Taussig, Henderson, & Mack, in press). Two EIWA subtests, Vocabulary and Similarities, and two WAIS-R subtests, Block Design and Digit Span, were included in the newly established Spanish language battery. Because subtests from both the EIWA and WAIS-R were used, we were then able to compare how Spanish-speaking elderly performed on subtests of both versions of the Wechsler test.

For purposes of comparison, we also examined the performance of an English-speaking elderly group on the same four measures, except that they were all taken from the WAIS-R and they were administered in English. Because the WAIS-R was primarily standardized on English-speaking non-Latino United States residents, we assumed that the performance of this group reflected no bias or untoward error. Therefore, we used the performance of the English-speaking as a standard; any difference between the English-speaking and Spanish-speaking elderly was thought to reflect a difference in how the two tests operate, a difference in the two groups' level of performance, or both. We hypothesized that the WAIS-R scale scores of the Spanish-speaking would be significantly less than the WAIS-R scale scores of the English-speaking. In addition, we hypothesized that the EIWA scale scores of the Spanish-speaking would be significantly higher than the WAIS-R scale scores of the English-speaking.

The following steps were taken to assess whether the hypothesized differences were likely to be a function of the hypothesized biases in the Wechsler tests or a function of valid group differences in their intellectual abilities. First, we examined whether there were any group differences on age and education and, when necessary, statistical controls were used. Second, raw scores and scale scores of the Wechsler subtests were analyzed. A consistent pattern of differences between the two language groups for both raw and scale scores would argue for valid group differences, whereas inconsistencies between the raw and scale scores could suggest possible biases in the tests' standardization norms. Third, two measures of global cognitive impairment were also taken: the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and the Activities of Daily Living (ADL; adapted from Lawton & Brady, 1969, and Blessed, Roth, & Roth, 1968) scale. Consistent language-group differences on the Wechsler subtests and on these two measures would argue for valid group differences; inconsistencies between the measures, particularly if only the Wechsler subtests demonstrate language group differences, would suggest biases in the Wechsler subtests. Finally, impaired and non-impaired samples were tested. Significant deviations from the expected pattern that the impaired elderly will function well below the average range and the nonimpaired elderly will function within the average range would suggest possible errors in using the Wechsler. We acknowledge that any one of these steps by themselves is limited. However, the convergence of the available data should contribute to deciphering whether any observed differences between the Spanish-speaking elderly and English-speaking elderly on the Wechsler subtests are likely to be a function of the tests or a function of the two groups' cognitive-intellectual abilities.

Method

Subjects

The Spanish-speaking participants were 17 female and 8 male older adults with dementia of the Alzheimer's type and 17 female and 5 male nonimpaired older adults. Together their mean age was 70.43 years (SD = 9.70). They were identified as Spanish-speaking if their native language and primary language were reported to be Spanish. All Spanish-speaking participants were Hispanics residing in the United States. The English-speaking participants were 13 women and 9 men with Alzheimer's Disease (AD) and 10 women and 12 men with no evidence of cognitive impairment. The mean age of all English-speaking subjects was 73.07 years (SD = 9.45), and they were all Anglo-Americans except for 2 African-Americans. The impaired patients met National Institute of Neurology and Communicative Disease and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable Alzheimer's Disease (McKhann, Drachman, Folstein, Katzman, Price, and Stadlan, 1984). The nonimpaired (NI) were those with no known neurological disorders as indicated by their self-reported medical history, by their primary physician's report, and by a brief neurological assessment.

Procedures

The participants were recruited through the media (radio, television, and newspapers) and through community presentations in health, mental health, and senior citizen centers. The Spanish-speaking elders were recruited through the Spanish-speaking Alzheimer's Disease Research Program (SSADRP), and the English-speaking elders were recruited through the Alzheimer's Disease Research Center (ADRC). All participants received neuropsychological testing and a systematic intake interview regarding demographic background (e.g., age and educational level) and historical information about their level of functioning. In addition, NI elderly received a brief neurological screening test conducted by a registered nurse, and AD patients received a neurological exam and an assessment of their activities of daily living. For the AD patients, significant family members, companions, or friends served as informants. Clinical psychologists, doctoral students in clinical psychology, and trained assistants conducted the testing and interviews. The examiners who assessed the Spanish-speaking were bilingual.

Measures

Four Wechsler subtests constituted part of the neuropsychological battery and were selected for this study: Similarities, Vocabulary, Digit Span and Block Design. As noted in Taussig et al. (in press), these subtests were chosen because they measure specific areas of function-
ing related to AD; for example, the Digit Span subtest measures memory and is related to auditory recall, attention, and freedom from distractibility (Lindemann & Matarazzo, 1984). The Spanish-language version of the entire test battery was developed to be a close equivalent of the English-language version; whenever possible, identical tests were used. This is the case for the Digit Span and Block Design subtests; they are the identical subtests for the Spanish and English versions of the battery. The content, administration, and scoring were taken from the WAIS-R. The only difference between the two tests is the language in which the test is administered. In contrast, the Similarities and Vocabulary subtests are different for the two language versions of the test battery. The Spanish version was taken from the EIWA, and the English version was taken from the WAIS-R. Tausig et al. (in press) chose these two subtests from different tests despite their significant differences in content and scoring because selecting one version would have required translation of the items from one language to another. Translation of the stimulus words for the Similarities and Vocabulary subtests would not have resulted in conceptually and functionally equivalent items (Brislin, 1980; Marin & Marin, 1982).

It is important to note that the scale scores were derived from the general norms for both the EIWA and WAIS-R. The age-specific norms available for the WAIS-R were not used because the EIWA does not have age-scaled norms.

Two other measures included in the neuropsychological battery are the MMSE and the ADL. The MMSE was chosen because it is a widely used, reliable, and valid measurement of global cognitive impairment in the general population, and it has been shown to have reasonable validity and reliability among a Puerto Rican community sample, particularly when educational level is taken into account (Bird, Canino, Stipec, & Shrout, 1987). Although the range of scores generated from the original instrument is 0 to 30, for this and all research at the University of Southern California's ADRC and SSADR, the range of scores has been extended to a maximum of 35 points. The serial subtraction task and the reverse spelling task are included in the scoring, resulting in a total possible score of 35. The original MMSE permits scoring only one of these two items.

The ADL scale is a measure of behavioral impairment. It is an 18-item scale designed to assess the severity of an individual's level of behavioral impairment such as personal care, household care, and management of finances. The items are derived from a combination of behaviors represented in the Instrumental Activities of Daily Living Scale (Lawton & Brody, 1969) and the Blessed Dementia Scale (Blessed, Roth, & Roth, 1968). By combining items from both of these scales, a comprehensive assessment of behavioral impairment is obtained. Each item contains a range of scores from 0, no impairment, to 3, great impairment, resulting in a potential maximum score of 54. However, not all the listed activities are part of each elder's life style. For example, some older adults have other people prepare their meals even though they are capable of doing so. In cases such as these, to guard against ascribing impairment when none existed, we deleted items from the 18-item impairment rating scale that did not pertain to a given individual. To derive comparable scores for all participants, particularly for those in which one or more items were deleted, we computed a percentage score for each elder: the sum of the actual impairment ratings divided by the sum of potential number of impairment ratings one could receive. Thus, a higher percentage reflected more impairment.

Both the MMSE and ADL scales were translated using a back translation method in which the English version was first translated into Spanish, and then the Spanish version was translated into English (see Brislin, 1980). Furthermore, as outlined in Tausig et al. (in press), the back translators included bilingual psychologists from different Spanish-speaking countries of origin (Argentina, Colombia, Cuba, Mexico, and Puerto Rico) to ensure that the Spanish be understandable to the broadest range of Spanish-speaking individuals.

Results

To assess whether the age and educational level were different for the groups, 2 (Language Background: Spanish-speaking [SS] and English-speaking [ES]) X 2 (cognitive impairment: Alzheimer's Disease [AD] patients versus nonimpaired [NI] elderly) analyses of variance (ANOVA) were conducted. Regarding age, a main effect was found for the participants' cognitive impairment, F(1, 87) = 9.62, p < .005; AD patients were significantly older (M = 74.55, SD = 9.11) than NI subjects (M = 68.66, SD = 9.30). No main effect for the language background was found; the age of the two language groups was not significantly different (Spanish-speaking: M = 70.43, SD = 9.70; English-speaking: M = 73.07, SD = 9.45). With regard to educational level, main effects were revealed for Language Background, F(1, 87) = 38.80, p < .001, and Cognitive Impairment, F(1, 87) = 8.24, p = .005. The Spanish-speaking (M = 7.94, SD = 4.28) had fewer years of education than did the English-speaking (M = 13.07, SD = 3.68), as did the AD patients (M = 9.21, SD = 4.30) relative to the NI subjects (M = 11.70, SD = 4.90). No interactions were found to be significant. Given the potential confounds of age and educational level, the subsequent analyses were conducted using analyses of covariance (ANCOVA) with age and educational level serving as the covariates.

Raw Scores of Subtests by Language Background and Level of Cognitive Impairment

ANCOVAs (2 X 2) for the raw scores of Similarities, Vocabulary, Digit Span and Block Design reveal a consistent pattern of findings. For each subtest, there was a main effect for Cognitive Impairment (p < .001) and an interaction between Language Group and Impairment (p < .01). Only Digit Span and Block Design resulted in a main effect for Language Background (p < .05). An examination of the raw score means in Table 1 reveals that those with AD had significantly lower scores than did the unimpaired participants. Furthermore, multiple comparison tests (Tukey's Honestly Significant Difference tests) reveal that the interactions are due to Spanish-speaking nonimpaired elders having significantly lower scores (p < .05) than did English-speaking nonimpaired elders for each of the subtests. The means of Spanish-speaking and English-speaking AD patients were not significantly different (p < .05) for any of the subtests. Thus, the main effects for Language Background on the Digit Span and Block Design are primarily due to the differences between the English-speaking and Spanish-speaking participants without any known neurological disorder.

Scale Scores of Subtests by Language Background and Level of Cognitive Impairment

The analyses of the scale scores of the Similarities and Vocabulary subtests resulted in main effects for Language Background (p < .05) and for Cognitive Impairment (p < .001). A
significant (p < .05) Language Group × Impairment interaction was found only for the Vocabulary scale scores. An examination of the scale score means reveals a pattern that both parallels and deviates from the pattern observed with the raw score means. The parallel finding is that the scale scores of the AD patients are significantly lower than the scale scores of NI older adults, indicating that the AD patients' level of functioning is lower than the level of functioning of those without any neurological disorder.

The findings that deviate from the raw score pattern pertain to the Similarities and Vocabulary scale scores. Tukey's HSD tests reveal that for the Similarities subtest, the Spanish-speaking AD patients' scale scores are significantly greater (p < .05) than those for the English-speaking AD patients (see Table 1). The mean vocabulary scale scores are in the same direction, although the difference is not as reliable. A further difference from the raw score pattern is that no significant differences were found between the mean scale scores of the Spanish-speaking and English-speaking nonimpaired elderly on the Similarities and Vocabulary subtests. Thus, on the basis of the scale scores in which EIWA norms are used for the English-speaking and WAIS–R norms are used for the Spanish-speaking, the Spanish-speaking AD patients appear to be significantly less impaired than their English-speaking counterparts, despite the fact that their raw scores reveal no difference. Furthermore, for the nonimpaired participants, there appears to be no difference in the two language groups' level of functioning as assessed by scale scores, despite very significant differences in the two groups' raw scores.

For the Digit Span and Block Design subtests, significant main effects (p = .001) again resulted for the participants' cognitive impairment. Significant Language Group × Impairment interactions (p < .01) were also observed. Only for the Digit Span was a significant Language Group main effect found (p = .001). Multiple comparison tests and an examination of the means in Table 1 indicate that the pattern of findings are consistent with those observed with the raw score. Considering the effect of cognitive impairment, the means once again demonstrate that the AD patients performed at a significantly lower level than did the NI older adults. An examination of the mean scale scores by language background reveals that the Spanish-speaking nonimpaired means were significantly less than those for the English-speaking nonimpaired. The mean scale score differences for the Spanish-speaking and English-speaking Alzheimer's patients, however, did not reach significance (p < .05). Thus when using the identical subtests and the identical standardization norms generated from the WAIS–R, the Spanish-speaking AD patients do not differ from their English-speaking impaired counterparts; however, the Spanish-speaking nonimpaired elderly appear to perform at a lower level than the English-speaking nonimpaired elderly.

**MMSE and ADL Scores**

To examine the validity of the noted differences between participants' cognitive status and language background, ANCOVAs were conducted for the MMSE for all participants, and a t test was conducted for the ADL scores obtained for the AD patients only. A 2 (Language Background) × 2 (Cognitive Impairment) ANCOVA with education and age as covariates re-
Table 2
Means and Standard Deviations of Mini-Mental State Examination by Language Background and Cognitive Status

<table>
<thead>
<tr>
<th>Cognitive status</th>
<th>Spanish</th>
<th></th>
<th>English</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Alzheimer's patients</td>
<td>11.60</td>
<td>6.16</td>
<td>14.00</td>
<td>8.31</td>
</tr>
<tr>
<td>Nonimpaired elderly</td>
<td>27.64</td>
<td>4.78</td>
<td>31.27</td>
<td>6.33</td>
</tr>
</tbody>
</table>

Note. ns = 22 except for Spanish-speaking Alzheimer's patients (n = 25).

revealed a significant main effect for Cognitive Impairment, $F(1, 81) = 99.61, p < .001$. Language Group and its interaction with Cognitive Impairment proved to be nonsignificant ($F$s < .7). (See Table 2 for the unadjusted means and standard deviations). Language group differences also failed to reach significance for ADL scores, $t(45) = -0.23, p > .05$; Spanish-speaking and English-speaking AD patients had nearly identical scores ($M = .58, SD = .20; M = .56, SD = .23$, respectively). The results concerning the MMSE and ADL support the hypothesis that the AD patients and the NI participants were functioning at the same general cognitive level, regardless of their language background.

Partial Correlations of Wechsler Subtests With MMSE and ADL

One plausible explanation for the findings that there are no differences between the two language groups on the MMSE and the ADL, but that there are group differences on the Wechsler scales, is that the relationship among the three measures differs given the linguistic group. For example, the Spanish-language Wechsler subtests may be unrelated to the MMSE and ADL, whereas the English-language versions of these tests are related. To examine this possible explanation, separate second-order partial correlations were computed for the Wechsler scale scores as they relate to the MMSE and ADL for each linguistic group. Because age is correlated with the MMSE for all participants ($r = -.34, p = .001$) and with the ADL for AD patients ($r = .34, p = .02$), and because there are differences in the educational level of the Spanish-speaking and English-speaking participants, second-order partial correlations were computed controlling for age and education. An examination of Table 3 indicates that the Wechsler subtest scale scores correlate highly with scores obtained from the MMSE for the respective Spanish-speaking and English-speaking samples. There does not appear to be any significant difference in the relationship between the Wechsler subtest scale scores and the MMSE for the Spanish-speaking and English-speaking samples.

In contrast to the MMSE, only the Vocabulary and Block Design subtests were related to the ADL. Most important, this pattern of results is generally consistent for both the Spanish-speaking and English-speaking patients. The only notable difference is that the Block Design and the ADL is not significantly related for the Spanish-speaking patients but is significantly related for the English-speaking patients. Other than this difference, the relationship between the Wechsler scale scores and the ADL scores is relatively similar for both language groups.

Discussion

The available data support the hypotheses that there are biases in the cognitive-intellectual assessment of Spanish-speaking elderly when using the WAIS-R and the EIWA. Perhaps the most striking potential error occurs when using the WAIS-R subtests to assess nonimpaired Spanish-speaking older adults; the WAIS-R subtests indicate that they are functioning in the impaired direction (mean scale scores of 5.64 and 5.05). Further, the assessed level of functioning of the Spanish-speaking elders is significantly lower than that of the English-speaking elders (mean scale scores of 10.00 and 7.75). These findings are consistent with the hypothesis that using the etic norms of the WAIS-R can lead to underestimating the cognitive-intellectual level of Spanish-speaking adults.

The data also indicate that using the EIWA subtests can lead to an overestimation of Spanish-speaking elders' level of functioning. When the EIWA subtests are used to test the impaired

Table 3
Second Order Partial Correlations for Wechsler Scale Scores and Scores on the Mini-Mental State Examination and Activities of Daily Living Scale

<table>
<thead>
<tr>
<th>Wechsler scale</th>
<th>Scales</th>
<th>Similarities</th>
<th>Vocabulary</th>
<th>Digit span</th>
<th>Block design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini-Mental State Exam&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Spanish-speaking</td>
<td>.781***</td>
<td>.770***</td>
<td>.554***</td>
<td>.531***</td>
</tr>
<tr>
<td></td>
<td>English-speaking</td>
<td>.714***</td>
<td>.864***</td>
<td>.702***</td>
<td>.718***</td>
</tr>
<tr>
<td>Activities of Daily Living&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Spanish-speaking</td>
<td>-.273</td>
<td>-.602**</td>
<td>-.032</td>
<td>-.335</td>
</tr>
<tr>
<td></td>
<td>English-speaking</td>
<td>-.048</td>
<td>-.559*</td>
<td>-.136</td>
<td>-.553*</td>
</tr>
</tbody>
</table>

Note. Partial correlations were controlled for age and education.

<sup>a</sup> Higher score equals more impairment.

<sup>b</sup> Lower score equals more impairment.

* $p < .05$. ** $p < .01$. *** $p < .001$. 
Spanish-speaking elderly and when the WAIS-R subtests are used to test their English-speaking counterparts, the scale scores indicate that the Spanish-speaking are less impaired (5.08, 5.96) than the English speaking (2.35, 4.00). The mean scale scores for the Spanish-speaking AD patients still reflect impairment; however, it is a lesser degree of impairment than reflected by the mean scale scores of the English-speaking AD patients.

We interpret these language group differences as biases of the specific subtests rather than as valid differences in the two language groups' level of functioning, for several reasons. First, statistical controls were used in an effort to correct for educational and age differences between groups. Second, with regard to the overestimation bias of the EIWA, the raw scores of the two language groups did not differ, suggesting that they performed the subtests at the same level. However, when converting the raw scores into scale scores, a significant difference resulted, indicating that the Spanish-language standardization norms translate raw scores to be higher than do the WAIS-R norms. A third reason for interpreting the group differences as test bias is that the measures of global cognitive functioning, the MMSE and ADL, did not demonstrate language group differences as did the Wechsler subtests. Fourth, with regard to the underestimation bias of the WAIS-R, no evidence of impairment or history of possible impairment was noted among the nonimpaired group from both language groups. Finally, the findings of both types of bias were predicted, given the previous research of López and Romero (1988).

It is important to point out that any one of these additional pieces of evidence, by themselves, are limited in demonstrating that the two groups' levels of functioning are alike and that the observed differences are a function of the tests. For example, statistical controls are less effective than methodological controls, and the global impairment measures may not be as sensitive as the Wechsler subtests. However, it is the convergence of the available evidence that support the interpretation of an underestimation bias in using the WAIS-R subtests and an overestimation bias in using the EIWA subtests.

Although there is evidence of biases in using the Wechsler subtests to assess Spanish-speaking elderly, the biases do not seem to apply in all cases. No differences resulted in assessing the Spanish-speaking and English-speaking patients on the WAIS-R subtests. These findings suggest that both tests accurately reflect the level of functioning of the two language groups. This interpretation is supported by the fact that the only group difference occurred when using the EIWA norms, which likely overadjusted the patients' scale scores. The findings of no differences on the global adjustment ratings (MMSE and ADL), on the raw scores for both the EIWA and WAIS-R subtests, and on the scale scores of the WAIS-R subtests, all support the notion that the Spanish-speaking and English-speaking AD patients are functioning at approximately the same impaired level. Thus, the WAIS-R subtests may not result in a biased assessment of significantly impaired Spanish-speaking elderly.

Another plausible explanation for the lack of differences between the Spanish-speaking and English-speaking AD patients could be a floor effect. Given the AD patients' severe level of impairment, the available measures may not have been sensitive enough to identify possible group differences. Further research that uses more sensitive measures of impaired cognitive-intellectual functioning or that studies patients with mild dementia may be able to test this hypothesis.

Another case of possibly no test bias pertains to the Similarities and Vocabulary subtests for the Spanish-speaking subjects; there were no differences in the scale scores of these subtests for the two language groups. One interpretation of this finding is that the English-speaking and Spanish-speaking subjects are functioning at similar levels, and the EIWA subtests are not biased in assessing nonimpaired Spanish-speaking elderly. Support for this interpretation is available from the MMSE data; when controlling for education and age, the two language groups did not differ on the MMSE. Furthermore, a brief neurological exam failed to find any evidence of impairment in either group. Other data, however, run counter to this interpretation. It is clear that the Spanish-speaking nonpatient older adults performed at a lower level than did the English-speaking nonpatient adults when examining all other Wechsler data, the EIWA raw scores and the WAIS-R raw scores and scale scores. Moreover, these differences remained after partialling out the effects of education and age. These findings suggest another interpretation of finding no scale score differences; that is, that the EIWA subtests are biased because their scale scores overadjust just the raw scores.

Although the interpretation of these findings is less than clear, we suggest that the culture-specific norms of the EIWA may be a more valid representation of the nonimpaired Spanish-speaking elderly's functioning. Compared with the WAIS-R standardization sample, the EIWA standardization sample more closely approximates the educational, language, and cultural background of the Spanish-speaking elderly sample. Therefore, given a choice between the EIWA norms and the WAIS-R norms for this relatively low-educated group of Spanish-speaking older adults, and given additional data to support the appropriateness of either set of norms, it is more likely that their performance will be best understood as it relates to the EIWA norms.

The variability of the findings (e.g., the EIWA in some cases may overestimate functioning and in other cases it may be an accurate measure) suggests that a strict adherence to either the emic norms of the EIWA or the etic norms of the WAIS-R will likely result in error. Given the considerable heterogeneity of Spanish-speaking elderly with regard to language background, educational level, cultural background, and level of impairment, we encourage psychologists to be flexible and to consider the specific circumstances of each patient or client. Some general guidelines, nonetheless, are offered to enhance the testing of Spanish-speaking adults.

The use of the WAIS-R or etic norms may be more appropriate when the testee speaks considerable English and is from a relatively higher educational or occupational background. However, with the WAIS-R, the examiner should be particularly cognizant of the potential bias of underestimating the functioning of less impaired elderly. In fact, the specific finding of the underestimation bias when using the WAIS-R subtests casts serious doubt on the practice of using the performance subtests of the WAIS-R, such as the Block Design, in testing Spanish-speaking adults. The assumption underlying this practice is that
there are little if any sociocultural biases in the test stimuli that could contribute to a lower level of performance. This may or may not be true. The important point here is that the norms are inappropriate for this population because monolingual Spanish-speaking persons were not included in the WAIS-R norms.

The use of the EIWA or culture-specific norms may be appropriate when the testee is monolingual Spanish and has a lower educational level. However, psychologists should be aware of the possibility that in using this test, particularly with significantly impaired elderly, they may overestimate their patient's level of functioning.

Culturally sensitive assessment then, requires a careful consideration of emic and etic norms for the specific individual being tested and the possible biases inherent in taking either approach. By including a good history, other tests, clinical observations, and the reports of significant others, the culturally sensitive psychologist can better determine whether the data obtained from the selected Wechsler subtests are the reflection of the person's actual level of functioning or possible test bias.

With the growing number of linguistic minorities in the United States it is important that the complexities in testing these populations be addressed (Olmedo, 1981). The present research is limited by the use of only four Wechsler subtests, the nonrepresentative samples of AD patients and nonimpaired elderly, and the lack of a clear "gold standard." Nevertheless, this research points out that testing the Spanish-speaking elderly with Wechsler scales is prone to significant error. Awareness of the potential errors in using either the WAIS-R or EIWA and the implementation of appropriate steps to guard against such errors should contribute to increasing the validity and utility in using the Wechsler scales to assess the cognitive-intellectual functioning of Spanish-speaking older adults.

References


Received July 3, 1990
Revision received November 30, 1990
Accepted January 18, 1991