Professional Psychology: Research and Practice
1988, Vol. 19, No. 3, 263-270

0735-7028/88/$00.75

Assessing the Intellectual Functioning of Spanish-Speaking Adults: Comparison of the EIWA and the WAIS

Steven López and Augusto Romero
University of Southern California

We point out the significant limitations in adapting the Wechsler Adult Intelligence Scale-Revised (WAIS-R) to assess the intellectual-cognitive functioning of Spanish-speaking adults, and we seek to familiarize practitioners with the Spanish version of the WAIS, the Escala de Inteligencia Wechsler Para Adultos (EIWA). To do the latter, we systematically examined both the EIWA and the WAIS and identified the exact differences between the two tests in regard to administration, content, scoring, and standardization sample characteristics. The most significant difference was found in the conversion of raw scores to scale scores. On several subtests, the equivalent raw score was converted into very different WAIS and EIWA scale scores. Other significant differences were noted in the content of the tests and in the social demographic makeup of the standardization samples. The administration and assignment of scores for both tests were generally found to be similar. On the basis of these findings, we offer specific recommendations for the testing of Spanish-speaking adults and for further research in this neglected area of study.

The growing population of linguistic minorities in the United States calls for a closer investigation of the appropriateness of using psychological instruments with these groups (Olmedo, 1981). Accordingly, there is an increasing number of researchers who are evaluating the reliability validity, and utility of Wechsler scales in the assessment of Hispanics, the largest linguistic minority group in this country. A close examination of this research reveals that it almost exclusively addresses the assessment of children. In a recent review of the literature, McShane and Cook (1985) identified more than 70 empirical studies in which the Wechsler scales for children were used and only two studies in which a Wechsler scale for adults was used (Murray, Waites, Veldman, & Heatly, 1973; Overall & Levin, 1978).

STEVEN LÓPEZ, PhD, is an assistant professor of psychology at the University of Southern California in Los Angeles. His main area of research concerns how patient variables such as ethnicity, gender, and age influence mental health professionals' clinical judgments.

AUGUSTO ROMERO received his MS in psychology from California State University in Los Angeles. He is currently employed as a registered psychological assistant at the Job Stress Clinic in Santa Ana, California, and he is enrolled in a doctoral program in clinical psychology at the Cambridge Graduate School of Professional Psychology in Los Angeles. He is also Vice President of the California Hispanic Psychological Association.

THIS ARTICLE WAS PREPARED while the first author was a recipient of a Ford Foundation Postdoctoral Fellowship for Minorities. During this time he was affiliated with the Spanish Speaking Psychosocial Clinic of the Neuropsychiatric Institute, University of California, Los Angeles.

WE THANK Frank Acosta, Robert Lewis, and anonymous reviewers for their helpful comments, and we thank Joseph Núñez for his ratings, used in the reliability check.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Steven López, Department of Psychology, University of Southern California, Los Angeles, California 90089-1061.

Although McShane and Cook's (1985) review was very comprehensive, they did not include a handful of studies in which researchers examined Hispanic adults' performance on Wechsler intelligence tests. In two such studies, the researchers assessed the relationship between the Wechsler Adult Intelligence Scale (WAIS) or the WAIS-Revised (WAIS-R) and shorter forms of intelligence tests, the Satz-Mogel short form (Adams, Kobos, & Preston, 1977), and the Revised Beta Examination (Hiltonsmith, Hayman, & Kleinman, 1984). In two other investigations, Green (1969, 1972) examined the relationship of (a) intelligence and (b) the age and race of Puerto Ricans. Kunce and Schmidt del Vales (1986) derived the factor structure of the subtest scores on the Spanish WAIS for a sample of Mexico City workers. In a final study, one that is particularly relevant to our study, Davis and Rodriguez (1979) investigated the comparability of the English and Spanish versions of the digit span and vocabulary subtests. To our knowledge, these eight investigations represent the only published empirical research (in English language journals) concerning the performance of Hispanic adults on Wechsler scales of intelligence. None of these investigators addressed the strengths and weaknesses of the English and Spanish Wechsler scales in the intellectual assessment of Spanish-speaking adults with limited English proficiency.

Current Practices

In testing Spanish-speaking adults' level of cognitive and intellectual functioning, psychologists should be using tests that have been developed and standardized for the population being tested. However, no Wechsler tests have been developed for United States residents who are predominantly Spanish speaking. As a result, practitioners who wish to use a Wechsler scale have two choices: adapt the English language test or use the Escala de Inteligencia Wechsler para Adultos (EIWA), a Spanish adaptation of the WAIS, standardized in 1965 with residents of Puerto Rico. Obviously, psychologists could

choose tests other than the Wechsler scales, or they could decide not to use standardized tests whatsoever. Our contacts with practicing psychologists, however, indicate that they are using standardized tests with Spanish-speaking adults, either some adaptation of the WAIS-R or the EIWA.

Some of the ways in which we have observed the WAIS-R used include (a) administering the instrument in English and attempting to take language difficulties into account when interpreting the scores, (b) administering only the performance subtests, using either English or Spanish instructions, (c) using an interpreter, or (d) referring the testing to a Spanish-speaking colleague or assistant who can translate instructions and test items during the test administration.

Adherence to any of these procedures is unsatisfactory and, in some cases, unethical. Should the test be administered in English, it is most difficult to ascertain the extent to which limited English language skills interfere with subjects' obtained scores, even for the performance subtests. Using only the performance section limits what can be said about a person's cognitive–intellectual abilities, especially if the examinee may not completely understand the examiner's instructions. Perhaps the least satisfactory alternative is to use any type of simultaneous or on-the-spot translation. As noted in the *Standards for Educational and Psychological Testing,* the reliability and validity of a test translation should be established before its use (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985). The practice of translating tests during their administration is invalid and unreliable and therefore should not be done.

The EIWA obviates many of the weaknesses associated with the WAIS-R. The instructions and items are in Spanish; therefore, the psychologist does not have to consider the influence of limited English language proficiency in interpreting scores. In addition, the content of the EIWA to some extent reflects the sociocultural background of Hispanics—in this case, Puerto Rican islanders. As a result, there should be fewer culturally biased items for Hispanics than there probably exist in the WAIS-R. Another noteworthy strength of the EIWA is that Spanish-speaking adults make up the standardization sample; therefore, the performance of a Spanish-speaking examinee can be compared with that of others from a similar linguistic background. Last, reliabilities and intercorrelations of scale and subtest scores have been computed and, in general, are within an acceptable range.

The EIWA is not without its limitations. In the Psychological Corporation's test catalog for 1986, the EIWA is described as *not* being interchangeable with the WAIS; however, the specific differences between the two tests are not noted. The EIWA test manual also provides little information about the two tests' comparability. The authors of the test manual generally stated that there are changes in the content, administration, scoring, and standardization. The only specific information provided is a table summarizing the number of EIWA items that are identical or different from WAIS items. In deciding between the EIWA and the WAIS-R, psychologists must make an informed decision; that is, they should know specifically how the EIWA differs from the English Wechsler scale and how these differences may affect the meaning of the obtained scores. Because of the lack of available information regarding the EIWA–WAIS comparability, professional psychologists may not even consider using the EIWA, or they may use it without full knowledge of its differences from the WAIS.

Study Objective

In order to familiarize psychologists with the EIWA and its comparability with the WAIS, we have systematically examined these instruments and have identified their specific differences in regard to administration, content, scoring, and standardization sample characteristics. On the basis of our findings, we then make recommendations for testing Spanish-speaking adults with limited English proficiency.

Method

Instruments

Escala de Inteligencia Wechsler para Adultos. The EIWA was standardized on 1,127 native Puerto Ricans of both sexes, aged 15–64. These subjects were selected in 1965 from the total population of the island on the basis of a comprehensive sample of residences provided by El Departamento del Trabajo (Department of Labor). Steps were taken to ensure that the sample accurately reflected the various geographical regions, occupational levels, the rural/urban nature of Puerto Rico, and other key demographic variables. The reliability of the performance, verbal, full scale, and subtest scores was assessed on the basis of the performance of three age groups. Reliability coefficients ranged from .65 to .96 for subtest scores and from .95 to .97 for performance and verbal scores. The reliability coefficient associated with the full scale score was .98.

Wechsler Adult Intelligence Scale. The WAIS is an updated version of the original 1939 Wechsler Bellevue Intelligence Scale. It was standardized on a national sample of 1,700 subjects of both sexes, aged 16–64, plus an additional 475 subjects of both sexes, aged 60 and older. The composition of the sample was based on figures obtained from the 1950 United States census. Efforts were made to obtain appropriate representation from sociodemographic groups according to gender, age, rural/urban community, geographic region, and other key variables. Although Wechsler (1955) indicated that the standardization sample included a proportionate representation of the non-White United States population, the actual size and composition of this subsample was not noted in the manual. Reliability coefficients ranged from .60 to .96 for the individual subtests and from .93 to .96 for the performance and verbal scores. The reliability of the full scale score was .97.

The WAIS has since been replaced by the Wechsler Adult Intelligence Scale-Revised (WAIS-R). This revision was conducted to update the content of the test and to develop norms that were based on contemporary population samples (Wechsler, 1981). We chose to compare the EIWA with the WAIS, instead of with the currently used WAIS-R, because the EIWA was derived from the WAIS. This comparison should then tell more about the EIWA itself. In contrast, an EIWA/WAIS-R comparison would reveal less about the EIWA because EIWA/WAIS-R differences would reflect WAIS/WAIS-R differences, in addition to English/Spanish Wechsler differences.

A related advantage in comparing the EIWA with the WAIS is that two of the subtests are identical with respect to content, administration, and scoring, thereby providing a means to assess performance differences between the two standardization samples that can not be attributable to test differences. Such comparisons cannot be made between the EIWA and the WAIS-R because there are no subtests that are identical in all respects. The subtests that are most similar across

the two tests are those of object assembly and digit span. Although their content is essentially identical for the EIWA and WAIS-R, the scoring does differ.

Procedure

The second author, Augusto Romero, a bilingual doctoral student in clinical psychology with considerable testing experience, carefully examined the manuals, scoring sheets, and materials for the EIWA and the WAIS. This examination was conducted for each subtest and followed the order presented in the manuals, which was the same for both tests. Romero simultaneously reviewed each EIWA and WAIS subtest and identified the differences in the two tests' administration (how each subtest is given, such as the number of failures needed before a subtest is discontinued), content (the items contained in each subtest), and scoring (the values assigned to subjects' responses and to the conversion of raw scores to scale scores). In addition, he identified the specific subtest items that were identical for the two tests.

To assess the reliability of this analysis, a second rater, a bilingual graduate student in clinical psychology with assessment experience, conducted the same review for three subtests. He was informed of the study's purpose after the completion of the ratings. As an additional test of the primary judge's accuracy in identifying content differences across *all* subtests, his ratings were compared with a similar analysis reported in the EIWA manual (Wechsler, 1968, p. 6). While making his ratings, the primary judge was unaware of the findings of the latter analysis.

To examine the conversion of raw scores to scale scores for the two tests, we estimated the means and standard deviations for the subtests. An examination of the actual means and standard deviations would have been best, but they have not been published by the Psychological Corporation. We calculated estimates of the subtest means by using the raw score associated with a scale score of 10. Given that one standard deviation equals three scale scores, we estimated the raw score standard deviations by calculating the differences between the raw scores (RS) that correspond to the means and the raw scores that correspond to the 13th and 7th scale scores (SS). We then divided the sum of these differences by 2 to obtain the mean raw score standard deviation. The following formula summarizes these calculations: [(RS of SS 13 − RS of SS 10) + (RS of SS 10 − RS of SS 7)]/2. When the raw score associated with a scale score was presented as an interval, the midpoint was used.

Once the estimates were calculated, EIWA raw scores were converted into WAIS raw score units if the total number of possible points for a given subtest were not equal for the two tests. For example, there are 32 and 29 possible points for the respective EIWA and WAIS information subtests. The estimated EIWA scores for this subtest were then multiplied by 29/32, or 0.906. This conversion ensured that the units used for both tests represented approximately the same level of performance.

A final comparison of the EIWA and WAIS concerned the standardization samples. The two samples were examined according to place of residence (rural or urban), occupation, and education. Rural communities were defined as cities of less than 2,500 habitants, as per Wechsler scales' definition.

Results

Reliability Check

The primary judge's difference ratings were compared with the independent ratings of a second judge for three randomly selected subtests: comprehension, arithmetic, and object assembly. The two judges agreed on all but 1 of the 31 EIWA/WAIS differences for these subtests, which reflected a 96.8% interrater agreement score. In regard to comprehension, they identified two administration differences, two items with scoring differences, and seven items with content differences. For arithmetic, no administration differences were noted, and six items differed in their scoring. The only discrepancy between the two raters occurred with the content of one item of the arithmetic subtest; thus they agreed on 13 of 14 different items. With respect to object assembly, both raters concurred that there were no differences between the two tests. On the basis of this reliability assessment, the difference ratings of the primary judge appear to be highly reliable. In fact, the interrater agreement score actually underestimates the agreement between raters, given that judgments of identical administration, scoring, and content were not included in this analysis.

To assess the reliability of the judge's ratings on content similarity for all subtests, we calculated a second interrater agreement score. In the EIWA manual, Wechsler (1968) reported the number of identical items for the WAIS and EIWA by subtest. The primary judge's ratings were then compared with the ratings presented in the manual. The actual interrater agreement score could not be computed because the individual item ratings of Wechsler (1968) were not reported. Therefore, only the reported sums of identical items for each subtest by both raters could be compared. This procedure could inflate the actual interrater agreement. Without both sets of ratings for each item, however, the amount of error in this interrater agreement score could not be determined.

In Column 2 of Table 1, the judgments of the two raters are presented for each subtest. A comparison of the number of identical items per subtest reveals that the two raters appear to have agreed on 78 of 82 items (95%). This relatively high level of agreement is also noteworthy when one considers that both raters identified the same number of identical items on 7 of 10 subtests. (The digit symbol subtest was excluded from this analysis because of the different subtest formats). There was only one apparent discrepancy in block design and arithmetic and only two apparent discrepancies in vocabulary. Although these figures appear to reflect good interrater reliability in judgments of identical items, it is important to keep in mind that this is an estimate of the percentage of agreement between the raters. For purposes of summarizing the data, we used the ratings reported by Wechsler (1968) because he and his collaborators were the originators of the EIWA and, therefore, would know exactly which items are identical across both instruments.

Administration

The subtests of both instruments are administered in the same order. The following five subtests have identical administration procedures: digit span, digit symbol, picture arrangement, object assembly, and arithmetic. Two different administration procedures are noted in the six other subtests. One difference is that the examiner begins with different item numbers for the information, comprehension, vocabulary, and block design subtests (e.g., WAIS Vocabulary Item 4 and

Table 1
Number and Proportion of WAIS Items in EIWA

Subtests	Number of items		Number of identical items		
	WAIS	EIWA	Wechsler (1968)	This study	Proportion[a]
Information	29	32	13	13	.41
Comprehension	14	15	8	8	.53
Arithmetic	14	16	2	3	.13
Similarities	13	15	5	5	.33
Digit Span	14	14	14	14	1.00
Vocabulary	40	40	18	20	.45
Digit Symbol[b]	90	110			
Picture Completion	21	25	4	4	.16
Block Design	10	10	7	8	.70
Picture Arrangement	8	10	3	3	.30
Object Assembly	4	4	4	4	1.00
Totals	257	291	78	82	.43[c]

Note. WAIS = Wechsler Adult Intelligence Scale; EIWA = Escala de Inteligencia Wechsler para Adultos.
[a] Number of identical items, as indentified by Wechsler (1968), divided by number of EIWA items. [b] Ratings were not made for the digit symbol because of the different formats. [c] Does not include digit symbol.

EIWA Vocabulary Item 1). The other difference is seen in the number of failures necessary before the subtest is discontinued. The WAIS/EIWA differences are noted in the information, comprehension, similarities, vocabulary, picture completion, and block design subtests. Except for picture completion, more failures are required for discontinuing the EIWA subtests than the WAIS subtests (e.g., after seven failures on the EIWA vocabulary subtest and after five on the same WAIS subtest). For the English version of the picture completion subtest, the examiner presents all items regardless of the examinee's success or failure, whereas for the Spanish version the examiner is instructed to discontinue after nine failures.

Content

The findings pertaining to the content of the EIWA and WAIS are summarized in Table 1 by each subtest. The three sets of columns respectively contain (a) the number of items in each subtest, (b) the number of items judged by the two raters to be identical, and (c) the proportion of EIWA items taken directly from the WAIS.

In general, the EIWA contains more items that differ from the WAIS than are identical to the WAIS. This is revealed in the overall proportion of identical items (43%) and in the fact that 6 of 10 subtests (excluding digit symbol) have proportions of less than 50% identical items. Only the digit span and object assembly subtests are identical in the two test versions. The test version differences in content are further evident in the digit symbol subtest. There are six symbols used in the EIWA, in comparison with nine symbols used in the WAIS. In addition, there are 16 practice items in the EIWA and 10 practice items in the WAIS. Altogether, the content of the EIWA appears to be largely different from the WAIS. It is important to note that although most EIWA items may be different from the WAIS, the degree of difference may vary considerably.

Scoring

Minor differences in the assignment of scores were noted in four subtests: comprehension, arithmetic, vocabulary, and picture arrangement. In the comprehension and vocabulary subtests, the first two or three items can receive a score of 0, 1, or 2 for the EIWA and only 0 or 2 for the WAIS. An additional scoring difference concerns the time limits for the arithmetic subtest. The WAIS contains four items with 15-s time limits, six items with 30-s limits, three items with 60-s limits, and one item with a 120-s time limit. In contrast, the EIWA has two more items with 15-s time limits, the same number of items with 30-s limits, one more item with a 60-s limit, and no items with a 120-s time limit. On the average, there are fewer seconds allowed per item for the EIWA subtest (M = 31.9) than for the WAIS subtest (M = 38.6). Another difference in the assignment of scores is evident in the time bonus points allocated for a small number of items in the arithmetic and picture arrangement subtests. The WAIS arithmetic subtest contains four items (Items 11–14) for which one can earn a bonus point by correctly answering the items in less than 10 or 20 s. Similarly, bonus points can be earned on Item 7 of the picture arrangement subtest for performances of less than 120 s. The EIWA has no corresponding item in either of these subtests with identical bonus points. Although there are several items with EIWA/WAIS differences in the assignment of scores, it is important to note that there are so many more items with no such differences.

An examination of the conversion of raw scores to scale scores reveals major scoring differences. This is most clearly seen in the relationship between raw scores and scaled scores for the two tests' digit span and object assembly. As noted earlier, these are the only subtests that are identical across the two language forms. In Figure 1 the raw scores for these subtests are plotted with their corresponding scale score. The EIWA curves are considerably more elevated than the WAIS curves for both subtests, indicating that a given raw score leads to a higher scale score on the EIWA than on the WAIS. For example, a raw score of 7 on the digit span subtest corresponds to a WAIS scale score of 4 and to an EIWA scale score of 9. Similarly, a raw score of 23 on the object assembly subtest equals a WAIS scale score of 7 and an EIWA scale score of 12. In both instances, there is a 5-point scale score difference for the identical raw scores.

Given the previously noted modifications made in the other nine subtests of the EIWA, it is difficult to determine whether the conversion differences noted in the digit span and object assembly subtests are also evident in the other subtests. In an attempt to test for possible conversion differences in the other subtests, we tentatively assumed that the level of difficulty for the EIWA and WAIS did not differ. We then estimated the means and standard deviations for each subtest of the EIWA and WAIS and conducted t tests to determine whether there were significant differences between the two standardization samples. Given the different curves depicted in Figure 1, we expected to find mean differences on the WAIS and EIWA's

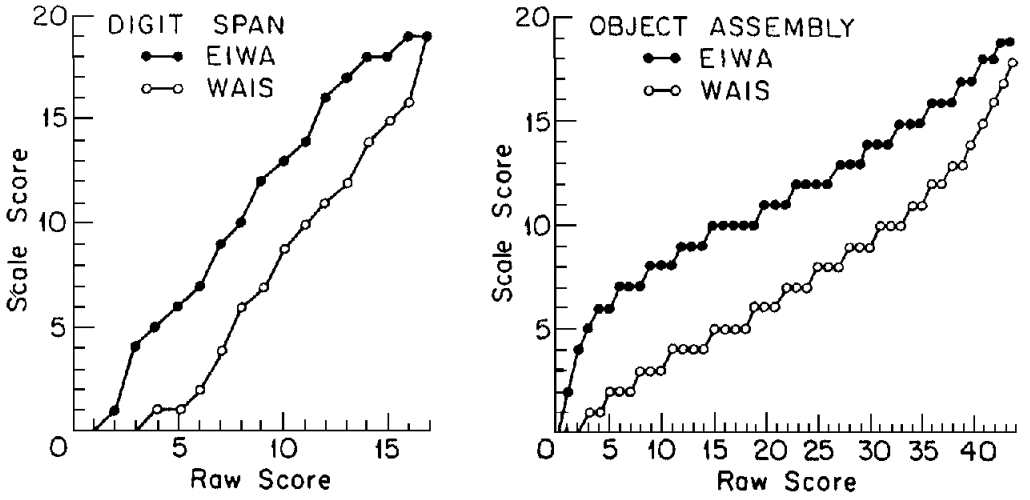

Figure 1. The relationship between raw scores and scale scores for the digit span and object assembly of the EIWA (Escala de Inteligencia Wechsler para Adultos) and the WAIS (Wechsler Adult Intelligence Scale).

digit span and object assembly subtests. By comparing the observed differences obtained for these two subtests with the mean differences for the other nine subtests, we assessed the representativeness of the digit span and object assembly's conversion differences.

In Table 2 we list the estimated means and standard deviations for the EIWA and WAIS subtests, as well as the *t* values obtained in assessing mean differences. An examination of this table indicates that there are major differences between the estimated means and standard deviations of the WAIS and EIWA subtests. Across all subtests, the estimated mean performance of the Puerto Rican sample was significantly lower than the mean performance of the United States sample. We found this pattern of results for both the identical subtests and the modified subtests, which suggests that the conversion differences noted in the two identical subtests are likely to be evident in the other nine subtests.

Table 2
Comparison of Estimated Means and Standard Deviations for All WAIS and EIWA Subtests

	WAIS (n = 500)		EIWA (n = 376)		
Subtest	*M*	*SD*	*M*	*SD*	*t*(874)
Information	15.50	6.00	10.87	6.34	11.02*
Comprehension	17.50	4.25	9.80	5.37	23.58*
Arithmetic	11.00	3.25	7.87	3.37	13.85*
Similarities	13.50	6.00	6.50	4.99	18.28*
Digit Span	11.00	2.25	8.00	2.00	20.41*
Vocabulary	43.00	18.50	26.50	16.00	13.77*
Digit Symbol	54.50	14.75	30.68	15.95	22.75*
Picture Completion	14.00	4.75	11.34	4.83	27.96*
Block Design	32.50	10.00	21.91	13.26	13.42*
Picture Arrangement	24.00	7.25	15.75	10.31	13.84*
Object Assembly	32.00	7.75	17.00	10.50	24.23*

Note. WAIS = Wechsler Adult Intelligence Scale; EIWA = Escala de Inteligencia Wechsler para Adultos.
* $p < .001$.

Selected Characteristics of Standardization Samples

We conducted chi-square analyses to assess the comparability of the two samples' rural/urban status, occupational level, and educational background. To conduct these analyses, we calculated estimates of the actual distributions on the basis of the percentages reported in the test manuals. The results indicate that the two standardization samples differ significantly across each of these variables: for rural/urban samples, $\chi^2(1, N = 2{,}827) = 88.14$, $p < .001$; for male occupational status, $\chi^2(10, N = 1{,}373) = 57.29$, $p < .001$; for female occupational status, $\chi^2(10, N = 1{,}454) = 49.79$, $p < .001$; and for educational level, $\chi^2(4, N = 2{,}827) = 165.72$, $p < .001$. An examination of the distributions pertaining to place of residence reveals that in relation to the WAIS sample, the EIWA sample included more residents from rural areas (52.6% vs. 34.8%) and fewer residents from urban areas (47.4% vs. 65.2%). In terms of occupational status, Table 3 indicates that fewer men from the EIWA sample were represented in the higher status occupations, such as farmers and farm administrators, and more men were represented in lower status categories, such as farm laborers and laborers. Although the majority of women in both the EIWA and WAIS samples were homemakers, those who were not reflected a distribution of occupations that was somewhat similar to that of the men. In comparison with women in the WAIS sample, women in the EIWA sample were less likely to be regularly employed (e.g., as office workers) and more likely to be incidental workers, unemployed, or disabled. Last, Table 4 reveals the sample differences in regard to educational level; the EIWA sample had many more subjects with 8 years or less of schooling and fewer subjects among the four higher educational levels than did the WAIS sample.

Table 3
Occupational Categories by Gender and Standardization Sample

Occupation	Men		Women	
	EIWA	WAIS[a]	EIWA	WAIS[a]
Professional, technical and kindred workers	3.8	5.4	3.8	4.0
Farmers and farm managers	3.1	7.5	0.3	0.3
Managers, officials, and proprietors except farm managers	4.8	7.7	0.3	1.3
Clerical, sales, and kindred workers	8.6	11.1	6.8	12.7
Craftsmen, foremen, and kindred workers	12.8	14.6	0.3	0.2
Operatives and kindred workers	13.6	17.0	7.5	5.6
Private household workers and service workers	5.0	4.9	6.8	7.8
Farm laborers[b]	13.2	6.2	0.2	0.9
Laborers[b]	11.7	7.1	0.2	0.2
Keeping house	0.0	0.0	53.6	54.9
Students	13.2	12.5	13.6	10.7
Others[c]	10.3	5.7	6.6	1.5

Note. WAIS = Wechsler Adult Intelligence Scale; EIWA = Escala de Inteligencia Wechsler para Adultos. Because the percentages are rounded off, the column totals may not equal 100%.
[a] These figures are estimates derived from available data by Wechsler (1955).
[b] These occupation categories were collapsed for the chi-square analysis pertaining to women.
[c] Refers to those who are disabled, retired, unemployed, voluntarily idle, "incidental" workers (less than 15 hr a week), and so on.

Discussion

There are major differences between the EIWA and the WAIS. The most significant difference pertains to the conversion of raw scores to scale scores. As noted in Figure 1, there were several instances when the same raw score on the digit span and object assembly subtests would result in up to five scale scores that were higher on the EIWA than on the WAIS. If this discrepancy exists in several other subtests, then it is likely that similar full scale scores on each test reflect very different levels of performance. Presently, it is difficult to precisely determine whether these large discrepancies in raw-score/scale-score conversions go beyond the digit span and object assembly. The other subtests are not identical in nature; therefore, any noted conversion differences may reflect true differences in the level of difficulty between the two dissimilar subtests.

Table 4
Percentage of Subjects in Standardization Samples by Years of Education

Years of education	EIWA	WAIS[a]
8 or less	60.75	36.21
9–11	16.94	26.38
12	13.22	23.32
13–15	5.31	8.74
16 or more	3.73	5.34

Note. WAIS = Wechsler Adult Intelligence Scale; EIWA = Escala de Inteligencia Wechsler para Adultos.
[a] These figures are estimates derived from available data by Wechsler (1955).

The estimated means and standard deviations of the two standardization samples, however, are consistent with the hypothesis that this EIWA/WAIS discrepancy does exist in raw-score/scale-score conversions for most of the subtests. The conversion differences reflect what appear to be major differences in the performances of the United States and Puerto Rican standardization samples; the United States sample scored much higher than the Puerto Rican sample. This performance differential could explain why the identical raw scores led to such large differences in the scale scores. This interpretation, however, should be considered a hypothesis that requires further evidence. A more precise comparison of the two samples' performances should be carried out with the actual means and standard deviations for each subtest. Only then can we determine whether the two standardization samples differed as significantly as suggested by our analysis. An additional limitation of this specific analysis is that an equivalent level of difficulty for the two tests is assumed. This assumption has yet to be tested.

It is important to point out that the apparent EIWA/WAIS difference in scores does not mean that the Puerto Rican sample is less intelligent than the United States sample. An examination of the two samples' background variables reveals that subjects from Puerto Rico are more likely to reside in rural communities, to have lower status jobs, and to have less education than subjects from the United States. The apparent performance differences could very well be reflective of these important demographic differences. It is likely that a comparable United States sample would perform in a similar fashion as the Puerto Rican sample. In fact, analyses of the relationship between different levels of education and mean full scale IQ derived from the WAIS-R standardization sample show that persons with lower levels of education score significantly lower than persons with higher levels of education (Matarazzo & Herman, 1984). The purpose of comparing the performance of the two standardization samples is not to determine which group is more or less intelligent. Instead, we hope to point out why the WAIS and the EIWA differ. Understanding the reasons for the observed differences may assist the professional psychologist to judiciously use the tests.

In addition to the noted conversion and sample differences, we also found dissimilarities in the administration, content, and/or assignment of scores for all subtests except those of digit span and object assembly. The observed differences in the administration and assignment of scores appear to be of minor significance. They alone should contribute little to an examinee's level of performance. And as pointed out earlier, the impact of the content modifications on performance level, for the most part, is difficult to ascertain at this time. Only the EIWA digit symbol subtest seems to be considerably easier. Further research is needed to determine whether the content changes evident in the EIWA result in an instrument with a different level of difficulty than its predecessor, the WAIS.

Implications for Practice

Given this comparative analysis, we can make some statements regarding the use of the EIWA in assessing Spanish-

speaking adults with limited English proficiency. Psychologists should not expect the scores of the EIWA to be comparable with those of the WAIS, and perhaps even with the scores of the WAIS-R. The EIWA was standardized on a population very different from that of the WAIS. When the EIWA test results of a given subject are discussed, it is important that psychologists keep in mind the point of reference: Spanish-speaking individuals from predominantly rural communities, with little educational background (less than 9 years), and with lower status jobs. Accordingly, the EIWA may be more appropriately used with monolingual Spanish-speaking adults with this type of background. The test may be less appropriate for Spanish-speaking individuals who are highly educated, from urban areas, and who have higher status occupations.

Although this comparative analysis should be able to assist professionals in making more informed decisions about the use of the EIWA, and about the meaning of scores obtained from the EIWA, the decision to use the EIWA or the WAIS-R is still a difficult one to make. Both approaches have significant disadvantages. The clinician needs to carefully weigh the advantages and disadvantages that are relative to the purpose of the intellectual assessment. For example, if the pattern of subtest scores is more important than the actual full scale score, as it can be in neuropsychological testing, then maybe the EIWA would be the better instrument to use. By using the EIWA, the examiner can be assured that subtest variability is not reflective of limited English language skills.

If full scale scores are important, as they are in the assessment of mental retardation, the clinician has to make a judgment call. In using the WAIS-R, and not simultaneous translations thereof, the evaluator should be certain that a full scale score in the mental retardation range does not represent performance decrements attributable to limited language skills or to an unfamiliarity with the majority culture. To make this assessment, the psychologist should collect data from several sources, including behavioral observations, reports of significant others, a complete medical and developmental history, a thorough psychosocial history, and other test data. Similarly, it is important that evaluators do not dismiss findings of mental retardation as representative of limited English language skills and/or contact with the majority culture, without corroborating evidence. In adjusting for perceived language and cultural biases in the WAIS-R, the examiner is at risk for minimizing or normalizing performance decrements reflective of retardation.

Should the evaluator choose to use the EIWA, he or she should consider the possibility that the full scale score may be inflated, particularly if the individual is from a more educated background. For example, an adult with some college education who scores in the low 90s on the EIWA may be more impaired than the score actually suggests. Again, the Wechsler score should not be used in isolation; other data are needed to more definitively determine the accuracy of the EIWA score.

We offer the following specific recommendations to guide psychologists in their testing of Spanish-speaking adults. First, we recommend that evaluators fluent in Spanish be used whenever possible. Second, psychologists who use only the WAIS-R or some modification thereof should begin using the EIWA, especially for patients who may have language impairments and for patients from rural communities who have little formal education. Third, practitioners who insist on using the WAIS-R should be made aware of the significant limitations in using an English-language test for Spanish-speaking adults and the potential negative consequences for the examinee (i.e., having test results that are neither valid nor reliable). It is particularly important that psychologists not use interpreters or any type of simultaneous translation for which the translated test's reliability and validity have not been established. Adhering to such procedures represents an unethical use of psychological tests. Fourth, psychologists who insist on using the WAIS-R should reconsider the potentially ethnocentric assumption that the English Wechsler norms are more reflective of true intelligence than are the EIWA norms. This is not meant to imply that the EIWA norms are better; rather, the assumption that the EIWA norms are of limited value because they deviate significantly from WAIS or WAIS-R norms should be reevaluated. Our fourth recommendation is that practitioners who use the EIWA be fully aware of the standardization sample's characteristics, and the significance of those characteristics in interpreting the client's performance. We also recommend that psychologists carefully communicate to users of test data the specific test that was used in assessing the Spanish-speaking client's level of cognitive functioning, and the normative data on which the interpretation of the results was based. There is some indication that these procedures are not always followed (Chavez & Gonzales-Singh, 1980). Last, no psychologist should administer a Wechsler test in isolation of other significant clinical and test data. There are significant limitations in using either the English or the Spanish Wechsler scale; however, we believe that an intelligent use of either test can provide valuable data about a client's level of intellectual–cognitive functioning.

Implications for Research

Given the lack of empirically based studies pertaining to the intellectual assessment of Hispanic adults, much research is needed. First, there is a great need for an adult intelligence test standardized on a contemporary sample of Spanish-speaking adults. The EIWA data were collected more than 20 years ago; the normative data and perhaps some content are likely to be outdated, especially considering the socioeconomic changes in Puerto Rico and in other parts of Latin America during this time period. The restandardization of the EIWA or the development of a new test altogether would be costly and difficult to carry out, particularly if efforts were made to include other Spanish-speaking populations in addition to Puerto Ricans. Such a test, however, would likely be a significant improvement over the current Wechsler tests used with the Spanish speaking.

Until a new test is developed, it is very important to examine current assessment practices with Spanish-speaking adults. Such research could serve to enhance the quality of psychological testing conducted with this population. For example, if such surveys revealed that psychologists are using procedures that have no established reliability and validity (e.g., administering the WAIS-R through a translator), then attention could be drawn to this problem and professional measures to correct such practices could be initiated.

Concurrently, investigators should begin to examine the standard lines of inquiry in psychological assessment; the psychometric properties of the EIWA and WAIS-R should be assessed when used with Spanish-speaking adults. One possible way to assess the validity of the English and Spanish Wechsler scales is to adopt the research strategy of Davis and Rodriguez (1979) and administer both tests to bilingual Spanish- and English-speaking adults who are equally competent in both languages. A second strategy might consist of systematically translating one of the test versions and administering it to a group of monolingual Spanish- or English-speaking adults. The scores obtained with either strategy could then be correlated with typical validity markers, such as performance on other intelligence tests, academic grades, and/or occupational performance. The intercorrelations of separate subtests for each test version could also be computed. Data obtained from these validity tests could be used to develop empirically based guidelines in adjusting the scores obtained with current tests, if adjustments are necessary.

In addition to assessing the psychometric properties of these instruments, it would be important to examine the relationship between Spanish and/or English language proficiency and performance on the Spanish and English Wechsler scales. This line of research could potentially assist practitioners in properly considering language factors in their assessment of Hispanics' cognitive functioning. In addition, this research could contribute to a related line of inquiry: bilingualism and intelligence.

Conclusion

Given the growing population of Spanish-speaking adults residing in the United States and what appears to be an increasing use of psychological services (López, 1981), it is most important that we evaluate the appropriateness of available psychological instruments for this population. The intellectual assessment of Hispanic adults is an open field of inquiry that has considerable potential to improve the quality of psychological evaluations conducted with members of the largest linguistic minority group in the United States.

References

Adams, R. L., Kobos, J. C., & Preston, J. (1977). Effect of racial–ethnic grouping, age, and IQ range on the validity of the Satz-Mogel short form of the Wechsler Adult Intelligence Scale. *Journal of Consulting and Clinical Psychology, 45,* 498–499.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Chavez, E. L., & Gonzales-Singh, E. (1980). Hispanic assessment: A case study. *Professional Psychology, 11,* 163–168.

Davis, T. M., & Rodriguez, V. L. (1979). Comparison of scores on the WAIS and its Puerto Rican counterpart, Escala de Inteligencia Wechsler para Adultos, in an institutionalized Latin American psychiatric population. *Journal of Consulting and Clinical Psychology, 47,* 181–182.

Green, R. F. (1969). Age–intelligence relationship between ages sixteen and sixty-four: A rising trend. *Developmental Psychology, 1,* 618–627.

Green, R. F. (1972). On the correlation between IQ and amount of "white" blood. *Proceedings of the 80th Annual Convention of the American Psychological Association, 7,* 285–286.

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.

Kunce, J. T., & Schmidt de Vales, E. (1986). Cross-cultural factor analytic similarity of Wechsler intelligence scores for Mexican adults. *Journal of Clinical Psychology, 42,* 165–169.

López, S. (1981). Mexican American usage of mental health facilities: Underutilization reconsidered. In A. Baron, Jr. (Ed.), *Explorations in Chicano psychology* (pp. 139–164). New York: Praeger.

Matarazzo, J. D., & Herman, D. O. (1984). Relationship of education and IQ in the WAIS-R standardization sample. *Journal of Consulting and Clinical Psychology, 52,* 631–634.

McShane, D., & Cook, V. J. (1985). Transcultural intellectual assessment: Performance by Hispanics on the Wechsler scales. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 737–785). New York: Wiley.

Murray, M. E., Waites, L., Veldman, D. J., & Heatly, M. D. (1973). Differences between WISC and WAIS scores in delinquent boys. *The Journal of Experimental Education, 42,* 68–72.

Olmedo, E. L. (1981). Testing linguistic minorities. *American Psychologist, 36,* 1078–1085.

Overall, J. E., & Levin, H. S. (1978). Correcting for cultural factors in evaluating intellectual deficit on the WAIS. *Journal of Clinical Psychology, 34,* 910–915.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale.* New York: The Psychological Corporation.

Wechsler, D. (1968). *Escala de Inteligencia Wechsler para Adultos.* New York: The Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised.* New York: The Psychological Corporation.

Received May 1, 1987
Revision received August 5, 1987
Accepted August 12, 1987 ■