The Spanish Version of the WAIS: Some Ethical Considerations

Fernando Melendez
San Diego, California

ABSTRACT

This commentary looks at the changes imposed upon the official (and only) Spanish translation of the Wechsler Adult Intelligence Scale available in the United States, and finds major discrepancies that are likely to distort the final IQ measures by an average increment of 20 points. The issue of professional ethics in the use of this instrument is raised and, because of these concerns, the author recommends that the Spanish translation of the WAIS be avoided in clinical assessments in the mainland United States, and that, if this test is used at all, it be done with extreme caution.

The dearth of standardized testing instruments with which to conduct examinations in Spanish has been noted and deplored in the recent literature (Geisinger, 1992; Gómez, Piedmont, & Fleming, 1992; Marín & Marín, 1991). Such a lack of instruments is regrettable given the increasing number of first and second generation Hispanics in this country whose limited or absent English requires that they be examined in Spanish. Complicating this picture is the availability of tests which purport to be "Spanish versions" of well known tests in English, and which are used, perhaps uncritically, by psychologists attempting to meet the needs of Hispanic minorities. One of these tests is the Spanish version of the WAIS (Green & Martínez, in D. Wechsler, 1968) which was standardized on a Puerto Rican population.

López and Romero (1988) have discussed some of the differences between the WAIS and its Spanish version, pointing out variations in the content of the scales, in the demographic makeup of the standardization samples, and especially in the marked difference in the conversion from raw scores to scaled scores between the two tests. They issued several cautionary warnings about the use of

Address for correspondence: Fernando Melendez, 2477 Congress Street, San Diego, CA, 92110, USA.

Accepted for publication: January 30, 1994.
the Spanish version of the WAIS, but they did not directly raise the questions of ethics that derive from using the Spanish version of the WAIS in assessing Hispanics. These authors tentatively and cautiously recommend the use of the Spanish WAIS when examining Spanish-speaking subjects.

The Escala de Inteligencia Wechsler para Adultos (EIWA) was standardized on a sample of 1,127 adults which reasonably conformed to the demographic characteristics of the 1960 Puerto Rico census. These characteristics were different from the WAIS sample in that the Puerto Rican group contained significantly more residents from rural areas, and included many more subjects with 8 years or less of schooling (López & Romero, 1988). The main features of the test and its statistical properties follow the pattern of the WAIS: it consists of 11 subtests, and it yields Verbal, Performance, and Full Scale IQs; it has a mean of 100 and a standard deviation of 15 in each of the age intervals of the sample. A factor analytical comparison between the WAIS and the EIWA (Gómez et al., 1992) supports the similarity of factorial structures between the two tests.

If the statistical and factorial aspects of these two tests are similar, the content, scoring, and consequences of taking one or the other test are dramatically different. One certainly would expect that a proper translation across cultural (as well as linguistic) domains would result in significant changes of the items of a test, especially those which are clearly cultural. But the changes found in the EIWA are so pervasive that they appear to exceed any reasonable cultural correction by altering not only the content of the tests, but their length, cut-off points, and scoring. All of these changes make the EIWA a more lenient test, even to the extent that some answers which are marked wrong in English are marked right in Spanish. There should be no “cultural” or any other reason for scoring incorrect answers as being correct. For instance, on the Information subtest (32 items in the EIWA, discontinue after 7 errors; 29 items in the WAIS, discontinue after 5 errors) if the subject answers in English that “dark clothes attract heat,” or that yeast causes dough to rise as a result of a “chemical reaction,” the score is 0. Exactly the same answers in Spanish are scored 1. Likewise, if an English speaker can name only two of “three types of blood vessels in the human body” the score is 0, but in the Spanish version the score is 1.

Some of the EIWA tests remain reasonably similar to the WAIS (Digit Span, Arithmetic, Object Assembly) but others are changed for no discernible cultural or linguistic reasons, apparently to make them easier: Digit Symbol, for instance, consists of only 6 model items, not the 9 used in the WAIS, and the squares in which the symbols are supposed to be entered are significantly larger than for the WAIS targets.

Some test items, of course, should have been changed in order to achieve cultural and linguistic parity between the two tests. All of the Vocabulary items and most of the Information and Comprehension questions fall into this category, and indeed they were changed in apparently appropriate ways. For instance, the question about naming four men who have been presidents of the United States since 1900 was eliminated, and a question about naming three languages which
are frequently used in the Americas was added (oddly, full credit is given for naming just two of these languages).

In order to further the similarities between the two tests, the developers established the Vocabulary subtest by conducting a separate study from which they selected a list of words that had a similar gradient of difficulty (albeit for a less well-educated population) as the WAIS list.

In computing the IQ values, the EIWA follows the same clerical pattern of determining the scaled score for each raw score, adding the scaled scores of the six Verbal and the five of Performance domains, and then entering these values in the appropriate table for the age group of the individual. Nothing in the EIWA manual warns the user that the conversion from raw to scaled points is significantly different from the WAIS. Some differences in the conversion between raw and scaled points should be expected as part of the statistical translation from WAIS to EIWA, if the latter is to have the same statistical properties as the WAIS; but the translation project directors (Green & Martínez, 1968) should have warned the users of this test that the conversion from raw scores to scaled scores would diverge significantly from the WAIS figures. An illustration of the issue is shown below:

Since only two of the WAIS subtests (Digit Span and Object Assembly) were converted to the EIWA without any changes whatever, they will be used as an example. The stimulus, the rules for administration, and the scoring rules remained the same whether the tests were given in Spanish or in English.

Numbers do not change their symbolic characteristics when they are translated from one language to another, and therefore Digit Span can be assumed to remain the same task, whether it is conducted in Spanish or English. Similarly, the act of solving jigsaw puzzles has little linguistic import, although it may have some minor cultural component.

If someone is able to repeat six digits forward and five digits backward in English, he or she will obtain a scaled score of 10; if, however, that person can repeat the same six digits forward and five backward in Spanish, he or she would obtain a scaled score of 14. If someone is taking the test in Spanish and obtains 15 raw points on Object Assembly, the corresponding scaled score will be 10 points; but if the identical responses are made in “English,” the performance will only obtain 5 scaled points (Table 1).

This pattern of the EIWA’s inflated scores in relation to the WAIS extends to all the subtests. The effects of such inflation are very dramatic. For instance, suppose that a 23-year-old male correctly answers the first five questions of Information; the first four of Comprehension, with 2-point answers; the first four Arithmetic problems; the first three Similarities questions, with 2-point answers; that he repeats four digits forward and three backward; and he correctly answers the first nine Vocabulary words with 2-point answers. His Verbal IQ, in English, would be 71, or in the borderline mentally retarded range. If, however, he produced exactly the same results in Spanish, his verbal IQ would be 17 points higher, in the low average range of 88. Let us assume that this same person
obtained the following raw Performance scores: Digit Symbol 18, Picture Completion 4, Block Design 10, Picture Arrangement 8, and Object Assembly 11. His English language Performance IQ would be 60, and his Spanish IQ would be 82, or 22 points higher. The Full Scale IQs would be: English = 64 (mild mental retardation); Spanish = 84 (low average range).

In the middle range, the EIWA overestimates Full Scale IQs by about 20 points: the raw points needed to obtain a Full Scale IQ of 100 in a 23-year-old yields an IQ of 80 using the WAIS tables. In the lower ranges, the Full Scale IQs are also overestimated by about 20 points: the raw scores needed to obtain an IQ of 64 on the EIWA produce an IQ of 44 on the WAIS. In the upper ranges, the EIWA overestimates IQs by about 12 points: the raw points needed to obtain an EIWA Full Scale IQ of 124 produce a WAIS Full Scale IQ of 112.

These discrepant figures between the WAIS and the EIWA are considerably less than those previously reported by Davis and Rodríguez (1979), who obtained differences as high as 27 on the Full Scale, 25 on the Verbal, and 40 on the Performance scores for their sample, with the EIWA always yielding the higher scores; their study, however, was limited to the Vocabulary and the Block Design subtests of the WAIS and the EIWA.
López and Romero (1988) recommend that the EIWA be used with Hispanic populations in the U.S.A., provided that psychologists carefully communicate the normative data on which the interpretation of the results was based. Realistically, however, consumers of clinical information are likely to latch on to the summary IQ numbers found in a report, and pay scant attention to the warnings and cautions expressed about the meaning and interpretation of those numbers.

The EIWA and the consequent distortions it might spawn raise compelling ethical questions. The use of this test may push out of the reach of social services and other possible benefits a large segment of Spanish-speaking adults that would be eligible for such services if their IQs were correctly computed. Just as it is a violation of ethics to label people “retarded” on the basis of test results which are clearly culturally or linguistically biased against them (American Psychological Association, 1985), it is equally unethical to label individuals as having average or low average IQs on the basis of tests which distort their scores upward, and, thus, deprive them of benefits and services to which they might be otherwise entitled.

Similar ethical issues have to do with neuropsychological assessments. Using the EIWA will produce inflated scores that might lead to labeling people who have sustained head injuries as functioning in the above average or superior ranges of intelligence, when in fact they may be suffering from clear cognitive deficits.

One can sympathize with those who handled the restandardization and translation of the WAIS. The concept that a large segment of the Puerto Rican population had significantly lower IQs than the comparative sample of the WAIS must have been both scientifically and politically unsettling. However, one should not deal with differences in measurements by tampering with the yardstick; and if one does manipulate a measuring device in order to conform to preexisting ideals, it is only fair to explain how and why, and by how much, the dimensions were made different.

Perhaps the WAIS is a wholly inadequate instrument with which to measure the IQs of Puerto Ricans or other Hispanics, and an entirely different test needs to be devised that will more accurately reflect their intelligence; perhaps the demographic differences between the U.S.A. and the Puerto Rican samples resulted in extraordinarily different test performances; or perhaps there were significant differences between the IQ levels of the two societies from which the samples were drawn. Those and many other questions need to be addressed and resolved scientifically and impartially if the Hispanic population in the United States is to be properly served in the future.

Whatever the essential differences that may exist between the English-speaking and Spanish-speaking populations, however, it is both demeaning and patronizing towards Hispanics to use a test in translation which artificially boosts the IQ results of Hispanics in comparison to the English-speaking standards. The transformations caused by the EIWA may be an egalitarian’s dream, but they can also be a clinician’s nightmare; furthermore, the discrepancies of results that are
obtained between the use of the WAIS and its "Spanish version," give rise to serious ethical concerns. For this reason the author advises other clinicians to avoid using the EIWA in the mainland United States, if at all possible; and in those cases in which it is a choice between administering the EIWA or nothing at all, to use it, but to write their reports with sufficient caveats that the reader will not be lulled by a false sense of security into thinking that because the two tests are "Wechsler" the results they yield must be somehow equivalent.

REFERENCES


