A VALIDATED TRANSLATION INTO SPANISH OF THE WISC-R VOCABULARY SUBTEST WORDS

JOSÉ M. TAMAYO
Bureau of Child Study
Board of Education, Chicago, Illinois

The WISC-R vocabulary words were translated into Spanish utilizing the frequency of use of the words as a measure of their difficulty. The procedure was validated with a sample of 120 English and Spanish-speaking students (Mexicans, Puerto Ricans, and a group of other Central and South American nationalities). Both boys and girls were included in the study. Comparable results were obtained with both linguistic samples, suggesting that frequency of use is an adequate measure of word difficulty and that it is possible to validly translate vocabulary tests. The Spanish words were rearranged to make the passing percentage of the words as similar as possible in both lists.

Administering tests constructed in English in American schools to non-English-speaking children in their native languages is a necessity because of the limited availability of tests constructed in languages other than English that have been normed with bilingual American children. Given this circumstance, a validated translation of English tests is mandatory before a translated test can be used with non-English-speaking students. The term validated translation refers to (a) an equivalent content and level of difficulty of the items in both versions and (b) the attainment of similar results with comparable samples of English and non-English-speaking subjects. Tamayo (1985, 1987) demonstrated the viability of validly translating English vocabulary tests into Spanish vocabulary tests by considering the content or the denotative meaning of words and their level of difficulty as measured by frequency of use. The results of these investigations supported the validity of frequency of use as a
measure of word difficulty. The original study (Tamayo, 1985), however, utilized a list of words arbitrarily selected rather than a list of words taken from published tests.

The purpose of the present study was to investigate the concurrent validity of vocabulary scores and frequency of use, based upon lists of English words taken from standardized tests and translated into Spanish following the validated procedures used by Tamayo (1985), for a sample of bilingual children receiving the Spanish version of the vocabulary test and for a monolingual (English only) sample receiving the English version.

The Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) is the most frequently used test in American schools to assess the intellectual functioning of children referred for psychological evaluation. A comparable translation into Spanish of the WISC-R is much needed because the Spanish-speaking population is rapidly increasing and good translations with adequate norms for Spanish-bilingual children are not yet widely available. This study translated the WISC-R vocabulary subtest utilizing the frequency of use of the words as a demonstrated valid measure of word difficulty and verifying whether similar results were obtained by comparable English- and Spanish-speaking samples.

Method

Selection and Translation of Words

The 32 English words that make up the vocabulary subtest of the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) were selected to investigate the relationship between the frequency of use and the obtained scores when the words were given to both English- and Spanish-speaking samples. Two word characteristics, content and frequency of use, were considered in the translation process. A description of each follows.

1. Content: The 32 WISC-R words were translated into 32 Spanish words of similar content or meaning, in order to make both lists of words as similar as possible in their denotative meaning (considered in this study to be the most important characteristic of a word). The translated Spanish words were those adopted by the Diccionario de la Lengua Española (1970), edited by the Real Academia Española, and serves as the official dictionary of the Spanish language. For example, for the word "Mantis" ("any of various related insects that hold their forelegs folded as if praying") (Guralnik and Friend, 1968) there was no similar Spanish word in the Diccionario. Thus, the Spanish word "Insecto" was adopted.

2. Frequency of use: The index of frequency of use for each English and Spanish word was calculated by using Eaton’s (1967) word frequency dictionary. Frequency of use was operationally defined as the index reported in Eaton’s dictionary. Although it was not possible to match all words by their frequency of use, the two word lists generated in English and Spanish for this study, yielded an average frequency of use that was not significantly different. The mean frequency of use for the English list of words was 5.33 ($SD = 2.68$) and 5.05 ($SD = 2.41$) for the Spanish list ($r = .44$, $p > .05$).

It was hypothesized that the mean of the total vocabulary scores for each of the two samples (English only and bilingual Spanish speakers) would not be statistically significant. (Hypothesis 1). Further, a negative relationship between the index of frequency of use (the lower the index, the more frequently the word is used) and the obtained mean score for the English and Spanish words was also hypothesized (Hypothesis 2).

Subjects

A total of 120 students were selected from several Catholic schools in Chicago. Sixty spoke only English and 60 spoke Spanish in addition to English. There were 30 boys and 30 girls in both linguistic groups. In the Spanish-speaking group there were 20 Mexicans, 20 Puerto Ricans, and 20 from other Hispanic nationalities. Ten boys and 10 girls comprised each ethnic group. To be classified as Mexican, Puerto Rican, or Other, both parents had to be born in Mexico, Puerto Rico, or in the same Hispanic country other than Mexico or Puerto Rico, respectively. Some of the students were born in Chicago and others were born in the country from which their parents emigrated. All participants, both English- and Spanish-speaking, were of the same low-middle social class. Subjects were matched by choosing pairs of children of the same age, sex, school, grade, and academic achievement as estimated by their teachers. Subjects ranged in age from 8 to 16 years. The English-speaking children spoke no other language. The Spanish-speaking children spoke both Spanish and English but Spanish was the primary language used in the home.

In addition to the previously stated two hypotheses, related to the frequency of word usage, two additional hypotheses were generated. First, there would be no statistically significant differences
among the total score means on the vocabulary test for the three Hispanic groups (Hypothesis 3). Second, there would be no statistically significant difference between the total score means on the vocabulary tests for boys and girls in the Spanish sample (Hypothesis 4).

**Procedure**

The 32 WISC-R English words were administered to English-speaking subjects and the 32 translated words were administered to the Spanish-speaking subjects. The directions from the Wechsler Manual (Wechsler, 1974), were followed with the exception that the test was not discontinued after five consecutive failures. In order to control for the "I don't know" effect, when the difficulty of the words increased, the difficulty was counterbalanced by dividing the list into 8 groups of 4 words each, so that the most difficult words occurred systematically in one of the four positions in each list. The students' responses were scored 2, 1, or 0 points according to the WISC-R manual criteria.

**Data Analysis**

A mean score for each word was initially calculated by averaging the obtained raw scores (2, 1, or 0) for subjects in (a) the English and Hispanic groups, (b) the three Hispanic groups, and (c) the Hispanic boys and girls groups. Means and standard deviations were also calculated for the total word scores for both English and Spanish versions of the vocabulary lists for both the English- and Spanish-speaking samples. In addition, means and standard deviations were calculated for boys and girls and for the three national Hispanic groups. A t test of the differences between means for the two lists and a one-way ANOVA, to compare the three Spanish groups, were calculated. Pearson product-moment correlations were computed to determine the extent of relationship between the frequency of use and the obtained mean scores for both, English and Spanish, lists.

**Results**

As shown in Table 1, the mean vocabulary scores in the lists were $M = 1.00$ for the English-speaking and $M = .95$ for the Spanish-speaking total samples. The difference was not statistically significant ($t = .29, p > .05$), supporting Hypothesis 1. The Pearson product-moment correlation coefficient between the index of fre-
TABLE 2
Spanish Words

1. Cuchillo  
2. Bicicleta  
3. Reloj  
4. Sombrero  
5. Paraguas  
6. Burro  
7. Alféizar  
8. Uña  
9. Pelegrino  
10. Jueg  
11. Vaivenes  
12. Ladron  
13. Juniar  
14. Insecto  
15. Contagioso  
16. Diamante  
17. Estorbo  
18. Forzar  
19. Prevenir  
20. Enmigar  
21. Fábula  
22. Comparario  
23. Estrofa  
24. Espionaje  
25. Rivalidad  
26. Recluir  
27. Esmienia  
28. Neodesal  
29. Calumnia  
30. Arrasar  
31. Inminente  
32. Dilatorio

Discussion

The results of this study, replicating the procedures of a previous study (Tamayo, 1987), supported the hypothesis that frequency of use was a valid and reliable measure of word difficulty. When the frequency of use index was not significantly different between the original and the translated list of words, comparable vocabulary score results were obtained with two similar samples. Thus, the measure appears to be valid for a student sample comprised of Mexican, Puerto Rican, and other Central and South American nationalities, as well as for males and females.

The study also demonstrated the viability of translating existing vocabulary tests into Spanish. Validated translations have the advantage that the same normative tables can be used for the English- and non-English-speaking groups in order to compare the linguistic minority students with their school mates. It has been claimed, without empirical support, that linguistic minority children should be compared exclusively with their ethnic, linguistic, and socio-economic groups rather than with the national norms. Although the use of national ethnic norms could be an adequate procedure when testing newcomers or to rule out mental retardation, other educational advantages may be questionable. Using different norms for different groups may create confusion because students could be identified as functioning at one level compared to their ethnic group and at a different level compared to their classmates. It would be virtually impossible to compare students of different ethnic backgrounds. These comparisons are needed in order to know where the students stand in comparison to the general population of students. The general population, not the ethnic group, will be the reference group at the college level and in the job market. Ethnic norms usually give significantly higher standard scores for the same raw score than do national norms. This is the case of the WISC adapted in Puerto Rico and of the Test de Vocabulario en Imágenes Peabody (TVIP) normed in Mexico and Puerto Rico (Dunn, Padilla, Lugo, and Dunn, 1986). (The Peabody results will be reported in a study that is in preparation). These issues are important and must be resolved by empirical research rather than by speculation.

The results of this investigation offer evidence in favor of the non-bias position of vocabulary test with linguistic minorities. If properly translated, vocabulary tests could be an adequate and non-biased estimate of the linguistic abilities of minority children.

A surprising result of this study was that boys scored higher than girls in a vocabulary test. Although girls usually score higher than boys on this type of test, no plausible explanation for this finding could be offered.

Finally, a possible limitation of this investigation includes the rather small sample of the three Hispanic groups (n = 20) and the gender groups (n = 30) and the fact that all subjects were of the same low socioeconomic level. The conclusions drawn in this study would be strengthened by a replication of these findings with a larger sample that includes students of different socioeconomic levels.

REFERENCES


