ABSTRACT

A review of the research reveals that standardized intelligence and achievement testing presents linguistic, cultural, and psychological difficulties for Spanish-speaking children in terms of such internal or intervening variables as the language of the administration of the test, the extent of the verbal factor in it, and the ethnic background of its administrator. At the core of the problem has been the IQ testing of Spanish-speaking children. There is a need of paramount importance to modify the use of present IQ instruments and to develop new specialized instruments that utilize the language and cultural background of Spanish-speaking children to facilitate the assessment of their academic abilities. As in the case of IQ testing, on the standardized achievement tests, a verbal factor appeared to militate against the optimal performance of Spanish-speaking children. Those subtests most dependent on English language skills generally resulted in poorest performance, indicating a handicap in language ability rather than in learning ability. (DB)
SPANISH-SPEAKING STUDENTS AND STANDARDIZED TESTS

By Perry Alan Zirkel

This article appeared in the June 1972 issue of the Urban Review and is being reproduced with the Author's permission.

Perry Alan Zirkel is assistant professor of education at the University of Hartford (Conn.), where he is in charge of Evaluation of Connecticut's Migratory Children's Program. Beginning July 1, he will direct a bilingual program for the Teachers Corps.
NOTES AND REFERENCES

1. Spence et al. found that Mexican-American children from homes where both English and Spanish were spoken scored significantly higher in both the WISC and S.B. than Mexican-American children from homes where only Spanish was spoken.

2. Kittell obtained similar findings by comparing the verbal and nonverbal IQ scores of monolingual vs. bilingual third graders who were from 15 different language backgrounds. However, he found the bilingual children to have significantly higher verbal and nonverbal results than monolingual children in grade 5.

3. Whether the directions are in a language at all may also make a difference. In a study involving children from various language backgrounds, but not including Spanish-speaking children, Pintner found the results on a nonverbal test with nonverbal directions to be higher than those on a nonverbal test with verbal directions.

4. Brodie pointed out that research has indicated a general lack of predictability of IQ tests for disadvantaged children, causing school authorities in several cities to discontinue their use. For a discussion of the decision to discontinue IQ testing in New York City, for example, see "Test Ban" in List of Studies.

5. Moreover, he found that neither IQ test was significantly related to grade point average for both Anglo and Spanish-American children.

6. Cooper found somewhat different results for bilingual children in Guam whose native language was Chamorro. He found that verbal IQ tests generally correlated higher with the results of the California Achievement Test than did nonverbal IQ tests.

7. Lerea and Kohut, on the other hand, found evidence that an association factor may have been an advantage of bilingual children, from Greek, Polish, and Norwegian backgrounds in performing a verbal learning task.

8. The author would like to thank Mr. José Luiz Hernandez of UCN's Bilingual Education Office for his invaluable insights and professional assistance in the development of this article.

9. Robinson similarly found the MAT to be comparably reliable for "advantaged", "average", and "disadvantaged" pupils, black and white.

10. G.F. Johnson questioned the content validity of the MAT for Title I pupils in general. "A middle-class oriented achievement test is validated by middle-class criteria. Thus, the middle-class culture bias is not eliminated from the tests."
on the Wechsler Intelligence Scale for Children (WISC) in Spanish than in English. Manuel found that the average IQ of a sample of 98 Mexican-American pupils on the Spanish version of the Stanford-Binet Intelligence Test (S-B) surpassed that of the English version.

Keston and Jimenez, who found the opposite to be the case with a sample of Mexican-American 4th graders, admitted, however, that their results were probably attributable to the relatively low Spanish language development of their sample and to the use of a Spanish version of the S-B that was developed in Spain and that was not modified to take into account differences in dialect and culture.

Despite the higher scores generally obtained by translating existing standardized IQ tests into Spanish with appropriate cultural modifications, there is evidence that such procedures may not totally solve the problem of effectively testing Spanish-speaking children. As early as 1927, Pintrier emphatically stated: "It is perfectly absurd to imagine that any real comparisons can be obtained by translating tests from one language to another." Roca described the efforts of the Department of Education in Puerto Rico to adapt and determine the norms of the WISC, S-B and Goodenough-Harris Draw-A-Man Test (S-A), making changes to allow for differences in vocabulary frequency, cultural conditions, and order of difficulty, yet the three adapted tests produced generally depressed IQ scores for Puerto Rican children relative to American norms. As Roca noted:

There is no doubt that no matter how well an intelligence scale is adapted from one culture to another, there are cultural differences which will make the children from the second culture score lower than those from the first.

Cote similarly found generally depressed mean IQ scores for Mexican-American children on several nonverbal intelligence instruments administered with the directions in Spanish.

Watson and Goodenough and Morris concurred that the search for a culture-free intelligence test is futile. Stablein, for example, found the Davis-Eells Test, an attempt at a culture-free test of intelligence, to be as discriminatory as other standardized measures between Anglo- and Spanish-American students. Nevertheless, the construction of specific intelligence tests for different cultural groups seems both possible and worthwhile. As Ramirez pointed out, such tests must be based on awareness of differences in the cognitive and incentive-motivational styles as well as the communication style of such students.

IQ Tests: Verbal v. Nonverbal That Spanish-speaking children face a language barrier that is built into standardized intelligence tests is further revealed in studies comparing their results on verbal and nonverbal subtests. Several such studies found the average performance scale score of Mexican-American children on the WISC to surpass their average verbal scale score. Darcy had similar results when comparing the scores of Puerto Rican children on the nonlanguage and verbal sections of the Pintrier General Ability Test.
Moreover, verbal and nonverbal IQ tests have been found to discriminate differentially between Anglo- and Spanish-American children. Altus noted that the average verbal WISC score of Anglo-American children significantly surpassed that of Mexican-American children, but that their respective nonverbal WISC scores did not differ significantly. Christiansen and Livermore recently replicated these results, in which, furthermore, the same pattern was reflected in the four related intellective factors of WISC described by Cohen. That is, they found Anglo-American students to significantly surpass Mexican-American students with respect to the two verbal factors (Verbal Comprehension and Relevancy), but not with respect to the two nonverbal factors (Perceptual Organization and Freedom From Distractibility). A consistently significant factor in the WISC performance of both the Anglo and Mexican-American pupils was socioeconomic status (SES).

Several studies confirmed Altus' results by employing other IQ instruments. For example, the same pattern emerged in B. E. Johnson's comparison of the results of Anglo and Mexican-American students on the verbal and nonverbal sections of the California Test of Mental Maturity. B. Johnson's sample of Mexican-American boys in grades 4-6 scored significantly below Anglo norms on a verbal IQ test, but their scores on a nonverbal instrument were not significantly different from the Anglo norms. Rice reported that Corwin arrived at similar findings by comparing the results of Anglo and Mexican-American children on various verbal and nonverbal intelligence tests. After administering a verbal IQ test to a group of Spanish-speaking youngsters, Coindray concluded that the instrument was actually a test of English vocabulary. Thus, verbal IQ tests obscure an adequate assessment of the mental ability of Spanish-speaking children.

Merely using a nonverbal IQ test, however, may be insufficient to assess accurately the mental ability of Spanish-speaking children. Some researchers have indicated that nonverbal and verbal IQ tests may measure different abilities. Further, other researchers have pointed out that so-called "nonverbal" tests contain a verbal factor. Finally, even the language of the directions of a nonverbal IQ test may make a difference. Whether this difference is significant or not remains a question. Pintner's view was that "nonverbal tests with verbal directions are not adequate. We cannot be sure translated verbal directions are equally hard or equally easy." According to Mahanian and Mitchell, scores of Spanish-speaking children in the primary grades on the Otis Group Intelligence Scale were significantly higher when administered with the directions in Spanish than with the directions in English. On the other hand, Anastasi and Cordero did not find the language of instructions to be a significant factor in their study of the use of the Cattell Culture Free Test with Spanish-speaking children. The Puerto Rican study similarly revealed no significant difference with regard to the performance of Spanish-speaking students on the Lorge-Thorndike Nonverbal Intelligence Test (L-T) with Spanish vs. English directions.
IQ Tests v. Achievement Criteria: Further indications of the questionable validity of IQ tests for Spanish-speaking children can be seen in studies exploring the relationship between their IQ scores and their results on measures of academic achievement. Such studies reveal differential results between Anglo and Spanish-American children. Moerer, for example, found that both the WISC and the LIT were significantly related to the results on the Metropolitan Achievement Test (MAT) for Anglo-American 9th graders, but that neither of these IQ measures was significantly related to the MAT for their Spanish-American counterparts. He reported that "the greatest differences between the Spanish-American and Anglo-ethnic groups were observed when reading ability and comprehension were most involved in the obtaining of measurement." Carrow similarly found differential effects of IQ on language achievement tests between Anglo and Spanish-speaking children. The importance of the language factor is even more evident in a study by Philippus which revealed that nonverbal IQ tests correlated higher with grade point average than did verbal IQ tests for Spanish-speaking students.

Mental Ability Test v. Tests: Studies by Jensen and by Rapier provide a further indication of the ineffectiveness of standardized IQ tests in determining the mental ability of Spanish-speaking children. Jensen devised a more direct method of measuring mental ability in the form of simple learning tasks. After testing a group of 36 Anglo and Mexican-American children equated on the basis of age, SES, and IQ, he found that Mexican-American children with low IQs performed not only significantly better than Anglo-American children with low IQs but also as well as both Anglo and Mexican-American high IQ children. Rapier also conducted two experiments involving various learning tasks and found evidence of different learning difficulties for Anglo and Mexican-American children who were matched on the basis of age, SES, and IQ, especially for those children in the lower range of IQ. She found a lack of verbal association to be one of the difficulties for Mexican-American children.

SUMMARY

A review of the research on the IQ testing of Spanish-speaking children reveals the linguistic and cultural handicaps standardized IQ tests present for such children. Such variables as the language, cultural construction, and extent of the verbal factor of such tests seem to significantly affect the performance of Spanish-speaking children. The validity of such tests is further called into question by experimenting with direct learning tasks and exploring the relationship to achievement criteria. Thus, there emerges the paramount need to modify the use of present IQ instruments and to develop new specialized instruments that utilize the language and cultural background of Spanish-speaking children to facilitate rather than obscure the assessment of their academic abilities.
A number of studies reported that Spanish-speaking children scored generally below Anglo-American children on standardized tests of academic achievement. Still, there is evidence that at least some of this discrepancy is caused by language factors.

Achievement Tests: Verbal v. Nonverbal. The discrepancy between Anglo and Spanish-American students was much greater in verbal than in nonverbal areas of measured achievement. For example, scores of verbal ability were consistently lower than nonverbal ability scores for Mexican-American and Puerto Rican pupils in grades 1 to 12. Palomares and Cummins noted that arithmetic subtest scores surpassed reading subtest scores of Mexican-American pupils in the early grades. Palomares and Johnson found that all but four of 50 Mexican-American pupils referred to ERM classes scored higher on the arithmetic subtest than on the reading and spelling subtests of the Wide Range Achievement Test. It is interesting to note that seven of these students were at or above grade level in all three achievement subtests and that only six of them had scores on the G-I.Q. indicating eligibility for ERM classes.

B.E. Johnson’s study is an example in itself. He administered two standardized achievement tests to a sample of 103 Anglo and Mexican-American 6th graders. He found that the Anglo subjects consistently surpassed the Mexican-American subjects in those subtests involving English language skills, but that there were no significant differences between the scores of the two groups on those subtests involving arithmetic skills.

Cline stated that socioeconomic as well as culturallinguistic factors should be considered in assessing the performance of Spanish-speaking students on standardized achievement tests. His sample of Anglo-American 7th graders appeared to outperform their Mexican-American counterparts on all subtests of the Stanford Achievement Test. However, with SES held constant at the lower level, these differences became insignificant except in the arithmetic subtest which favored the Mexican-American subsample.

Achievement Tests: Spanish v. English. Studies of standardized achievement tests which have been translated into Spanish offer more direct evidence of the language barrier that Spanish-speaking students face in such tests. When Mahakian, for example, administered a standardized reading test in both original and translated versions to 210 Spanish-speaking children in grades 1 through 7, 83 percent achieved higher total scores in Spanish, with decreasing scores found in ascending grades. Thonis similarly found that 16 out of 19 Mexican-American students scored higher when tested with a Spanish translation of the Peabody Picture Vocabulary Test than they did with the standardized English version. According to Davis and Personke the differences between Spanish and English administrations of the WRT were mostly nonsignificant for a group of Mexican-American 1st grade. However, the mean scores on the subtest that most appropriately reflected their language background (Word Meaning)
revealed a significant difference favoring the Spanish version. 

Despite repeated reminders of Gardner's statement at the 1967 U.S. Senate hearings on bilingual education, most writers and researchers on the subject of the education of the Spanish-speaking seem to have forgotten or neglected the importance of a major study conducted in Puerto Rico in 1926 by the International Institute of Teachers College, Columbia University, which involved the administration of over 69,000 standardized achievement tests in English and Spanish. The results of that study indicated that although English had been imposed as the language of instruction since the United States took control of Puerto Rico in 1893, the Puerto Rican children's achievement in English showed them to be markedly below that of continental children. But as Gardner emphasized, "The Puerto Rican children's achievement through Spanish was, by and large, markedly superior to that of continental children who were using their own mother tongue, English."

Translating standardized achievement tests points to, but does not provide, the way to more equitable opportunities for Spanish-speaking students. As Finch stated, "The development of tests appropriate to Spanish-speaking children is far more than simply translating existing tests." A glance at Eaton's cross-cultural frequency tests indicates the varying difficulty levels of lexical items across languages. Differences in dialect and spoken language further confound the intended equivalence of translated tests. The cultural boundaries that restrict meaning within languages are reflected in Hernandez' description of a commercially available Spanish translation of a standardized English achievement test. One of the items in the English edition called for the identification of the word pie by means of choosing the appropriate illustration from a series of pictures. In the Spanish edition, the item was translated, in accord with a "standard" Spanish-English dictionary, as pastel. The difficulty, however, became compounded rather than alleviated for Puerto Rican pupils, for pastel is a culturally and visually distinct dish for them in comparison with pie which has retained its "Anglo" verbal and visual identity in Puerto Rico.

Achievement Test: Essay vs. Objective. Caldwell and Mpwry provided further evidence of the importance of the language factor for Spanish-speaking children when they constructed objective and essay tests designed to be of the same content and difficulty in each. Despite the fact that the objective tests were given first, thereby causing any practice effects to accrue in favor of the essay tests, the Spanish-speaking students scored higher on the objective tests. L.W. Johnson similarly found that Spanish-speaking children scored significantly below their Anglo classmates on a test of English vocabulary, though the deficiency was less in subject matter vocabulary tests.
Achievement Tests: Reliability and Validity. Some researchers, particularly Fishman, have questioned the reliability and predictive validity of standardized achievement tests for minority group children in general. This question remains unsettled, although not unstudied, with specific reference to Spanish-speaking children. In their study of Spanish-speaking 1st graders, Mishra and Hurt found significantly lower levels of reliability and predictive validity for those subtests of the Metropolitan Readiness Tests (MRT) that were most dependent on English language ability. On the other hand, Mitchell considered the general level of predictive validity of the MRT for Spanish-speaking 1st graders to be comparable to that for other ethnic groups.

Despite different sample sizes, criterion, achievement tests, and lengths of the periods between testing, the reasons for these contradictory results are not totally evident and their clarification awaits further research. In a related study by Arnold, the MRT was quite reliable when used with Spanish-speaking 3rd graders, with the important proviso that an appropriate difficulty level be administered.

Where standardized readiness and achievement tests do prove to be reliable and valid predictors of later achievement test results, this may mean no more than confirming the consistency of language difficulties for Spanish-speaking children in school. Perssonke and Davis, for example, found the MRT to be a generally better predictor of reading ability for Spanish-speaking 1st graders when administered in English than when administered in Spanish. Their conclusion bears repeating:

perhaps the readiness test was valid but the reading program was not. Before accepting the program, or the test, as a valid predictor of success in that program, it might be pertinent to examine some of the alternatives.

SUMMARY

A review of the research on the use of standardized achievement tests with Spanish-speaking children reveals that, as in the case of the IQ testing of Spanish-speaking children, a verbal factor appeared to militate against their optimal performance. Those sub-tests most dependent on English language skills generally resulted in poorest performance, indicating a handicap in language ability rather than in learning ability.

PSYCHOLOGICAL REACTIONS TO TESTING

Many researchers have noted, but few have scientifically studied, the psychological reactions of Spanish-speaking students to testing. Palomares and Cummins indicated that Spanish-speaking children seemed to appreciate the special attention given to them through testing. But Armstrong and Smith noted a lack of test motivation in their Spanish-speaking subjects. Anastasi and Cordova described the characteristic reaction of their Spanish-speaking subjects to testing as "a mild confusion followed by amusement and indifference."
attributed this reaction to "linguistic bifurcation" between a Spanish-speaking home and an English-immersed school which resulted in "psychological insulation" to what goes on in the latter. As they further explained:

Not only test performance, but also the general intellectual development which tests are designed to gauge, are seriously handicapped by the attitudes and intellectual habits resulting from the child's early linguistic confusion.

Examiner Variable. The importance of linguistic and cultural background to the Spanish-speaking child is expressed in his reaction to the language and culture of the examiner. In a review of the state of the art of language testing for linguistically different learners, Borde noted the value of a test administrator's giving directions in line with the language background of the child.

The interplay of linguistic, cultural, and psychological factors is revealed in the examiner effects on the test performance of Spanish-speaking students. Anastasi and de Jesus described an enthusiastic response from their Spanish subjects to testing, which they attributed as the effect of an examiner of the same ethnic group as the students. Palomares and Johnson found indications of the effect of an examiner variable by comparing the referrals of Spanish-speaking children to PMR classes from Anglo- and Mexican-American school psychologists. Anastasi and Cordova also uncovered evidence of an examiner variable along the intertwined lines of language and culture by administering a nonverbal intelligence test to Puerto Rican pupils with directions in Spanish v. English. They found that a testing order beginning with Spanish favored the girls while one beginning with English favored the boys. They attributed this difference to the greater degree of acculturation of the boys and, thus, a greater rapport with the Anglo examiner.

The complexity of the linguistic, cultural, and psychological dynamics of the testing situation is further revealed in two recent studies involving examiner effects on the WISC performance of Spanish-speaking students. According to Swanson and DeLaissie, the use of a bilingual interpreter as an adjunct to a monolingual Anglo examiner did not significantly affect the WISC scores of a group of Mexican-American 1st graders. However, any linguistic advantages for such a testing situation may have been counterbalanced by the possibly intimidating presence of two adults. Moreover, the second-class status of these pupils native language may have been reinforced, rather than reversed, by the subservient position of the Spanish-speaking interpreter in relation to the Anglo examiner. A study by Thomas et al. showed that significantly different performance levels can be obtained for Spanish-speaking students on the WISC, depending on differences in examiner styles. Their study involved the testing of a group of Puerto Rican pupils by two Puerto Rican examiners. Although both examiners were equal with respect to sex, ethnicity, fluency in Spanish and English, and clinical experience, consistently higher scores were obtained by the examiner who encouraged active participation, verbalization, and repeated effort on the
part of the pupils. Their results suggest that such students may have more equitable opportunities on IQ and achievement tests if teaching and testing procedures are "optimized" rather than "standardized."

CONCLUSION

A review of the research reveals that standardized intelligence and achievement testing presents linguistic, cultural, and psychological difficulties for Spanish-speaking children in terms of such internal or intervening variables as the language of the administration of the test, the extent of the verbal factor in it, and the ethnic background of its administrator.

Researchers have indicated the inadequacy of many tests for Spanish-speaking children. Several of them (Hernández, Herr, Hughes and Sánchez, Rodríguez) pointed out that standardized tests do not take into consideration the nonstandardized background of Spanish-speaking students. In their "Guidelines for Testing Minority Group Children," Fishman et. al. underscored the need for developing different norms for specific minority groups (see also Rankin and Henderson). Because of such inadequacies, several writers (Dieppa, Flores, National Conference; Palomares and Cummins, Rice, Roca, Rodríguez, Willis) have decried the lack of appropriate tests and have called for the development of more effective instrumentation.

Bordie noted an emerging trend to develop new, specialized tests rather than to depend on traditional, commercially available instruments.

Rodríguez termed testing "an educational roadblock," which cuts Spanish-speaking children off at an early age from equal educational opportunity. He added that tests reflect the monocultural nature of the schools, not the bicultural background of such students. The "disadvantaged" label that is pinned on Spanish-speaking children by this monocultural school system is often cemented on by tests constructed from what for these children is a second language and culture. The recent developments of bilingual-bicultural education and tests, which use the linguistic and cultural background of the child as an asset rather than a liability, show that the label can be switched to "advantaged" Spanish-speaking children.12
NOTES AND REFERENCES

1. Spence et al. found that Mexican-American children from homes where both English and Spanish were spoken scored significantly higher in both the WISC and S.B. than Mexican-American children from homes where only Spanish was spoken.

2. Kittell obtained similar findings by comparing the verbal and nonverbal IQ scores of monolingual v. bilingual third graders who were from 15 different language backgrounds. However, he found the bilingual children to have significantly higher verbal and nonverbal results than monolingual children in grade 5.

3. Whether the directions are in a language at all may also make a difference. In a study involving children from various language backgrounds, but not including Spanish-speaking children, Pintner found the results on a nonverbal test with nonverbal directions to be higher than those on a nonverbal test with verbal directions.

4. Borkie pointed out that research has indicated a general lack of predictability of IQ tests for disadvantaged children, causing school authorities in several cities to discontinue their use. For a discussion of the decision to discontinue IQ testing in New York City, for example, see "Test Ban" in List of Studies.

5. Moreover, he found that neither IQ test was significantly related to grade point average for both Anglo and Spanish-American children.

6. Cooper found somewhat different results for bilingual children in Guam whose native language was Chamorro. He found that verbal IQ tests generally correlated higher with the results of the California Achievement Test than did nonverbal IQ tests.

7. Lerma and Kohut, on the other hand, found evidence that an association factor may have been an advantage of bilingual children, from Greek, Polish, and Norwegian backgrounds in performing a verbal learning task.

8. The author would like to thank Mr. Jose Luis Hernandez of USDOR's Bilingual Education Office for his invaluable insights and professional assistance in the development of this article.

9. Robinson similarly found the MAT to be comparably reliable for "advantaged," "average," and "disadvantaged" pupils, black and white.

10. G.F. Johnson questioned the content validity of the MAT for Title I pupils in general. "A middle-class-oriented achievement test is validated by middle-class criteria. Thus, the middle-class culture bias is not eliminated from the tests."
11. Zeboller questioned the applicability of the findings of the Armstrong study.

12. See e.g., Gates "Two-way" bilingual education programs can also facilitate the advantage of speaking Spanish for Anglo children. For a description of such a program, see Zirkel.

List of Studies Discussed in this Review:

- (Cohn, P. L. From the Editor *American Education*, 1970, 6.
- Caldwell, F. F., & Mowry, W. D. The essay versus the objective examination as measures of the achievement of bilingual children. *Journal of Educational Psychology*, 1933, 24, 696-702.
- Christiansen, T. A. (ed.) *Companion of Anglo ...


The "test-ban" in New York City. *Schools, Phi Delta Kappan, 1964, 46, 103-110.


