

DOCUMENT RESUME

ED 102 201

TN 004 258

AUTHOR Simon, Alan J.; Joiner, Lee M.
TITLE Adapting the Peabody Picture Vocabulary Test for Use
With Mexican Children.
PUB DATE [Apr 74]
NOTE 12p.; Paper presented at the Annual Meeting of the
American Educational Research Association (Chicago,
Illinois, April, 1974)

EDRS PRICE MF-\$0.76 HC-\$1.58 PLUS POSTAGE
DESCRIPTORS Culture Free Tests; Elementary Education; Elementary
School Students; *Intelligence Tests; Item Analysis;
*Mexicans; Statistical Analysis; *Test Bias; *Test
Reliability; Test Validity; *Translation
IDENTIFIERS *Peabody Picture Vocabulary Test

ABSTRACT

The effectiveness of test adaptation based on item selection and reordering of a Spanish (Mexican) version of the Peabody Picture Vocabulary Test (PPVT) was examined. Translated forms were administered to a sample of Mexican students. One item from each pair (A and B) was selected and reordered using a priori rules. The revised instrument was administered to a new cross-validation sample. Findings confirmed the cost effectiveness of this technique for improving reliability and validity over simple translation or the creation of completely new items for populations of different culture and language. (Author)

ED102201

Adapting The Peabody Picture Vocabulary Test
For Use With Mexican Children

Alan J. Simon, Ph.D.
Queens College, C.U.N.Y.

Lee M. Joiner, Ph.D.
Southern Illinois University

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
THE OFFICIAL POSITION OR POLICY OF
THE NATIONAL INSTITUTE OF
EDUCATION.

A Paper Presented at The Annual Convention
of The American Educational Research Association
Chicago, Illinois April 1974

TM004 258

Recent litigation involving children who were identified as mentally retarded by school systems has posed serious problems for educators and psychologists who routinely use tests for determining educational placement. A prime grievance of minority group parents has been that commonly used intelligence tests are unfair because of language and cultural biases. In the District Court of Northern California, a suit was filed on behalf of nine Mexican-American students whose primary language was Spanish but who were tested in English, and whose scores were sufficiently low (IQ 30-72) to warrant placement in classes for the mentally retarded. When retesting was conducted in Spanish, their average score increased 15 points and seven of the nine were above the mental retardation cutting point. On February 5, 1970 a stipulated agreement order was consummated which, among other things, required that children be tested in their primary language. Bilingual examiners and interpreters were to be used to implement this agreement (Weintraub and Abeson, 1974).

The problem that is encountered in attempting to implement decisions that mandate fairer testing procedures for Mexican-American children is that there is a scarcity of tests that are comparable across languages and cultures. Simple translations of achievement or intelligence tests result in tests having lower reliability, validity, means, and standard deviations.

Yet an examination of a collection of currently available tests (ETS, Test Collection, 1972) reveals that in almost every case, researchers have chosen to modify existing tests rather than become involved in the

costly highly technical and very time consuming task of developing new tests. The majority of these modifications take the form of simple translations. The Psychological Corporation catalogue is offered as but one example. It identifies nine tests on its list of "Spanish-Language Edition of Tests". Of these, seven were simply translated, and two, the WISC and the WAIS were "adapted to the Spanish Culture". In the U. S., researchers have often simply translated tests for use with non-English speaking children (Rieber, 1968).

After studying intelligence tests translated from English to Spanish as spoken in Puerto Rico, Roca (1955) concluded that one of the main problems in translating tests was that a translated word, although expressing identical concepts, may be of a different degree of difficulty in the new than in the old language. Hyness (1970) noted the problem that a concept or word may not be present in the new culture; the translation, then is accomplished by replacing the intended concept with one which is judged to be similar, this judgement being subject to debate. Another problem is that a word may possess a single meaning in one culture but possess multiple meanings in the other; or a word may have opposite meanings.

If adequate translation of tests is difficult to accomplish, neither will more extensive adaptations be free of problems. In this respect the importance of Ali's (1967) findings concerning the use of an American achievement test in East Pakistan can be seen. Item analysis showed that the translated version of the tests made a greater contribution to the research than did the adapted version. Adaptation

of test items in terms of time, work, and statistical results did not seem worth pursuing as a means of developing tests until and unless the criteria of and the elements for adaptation were determined through further experimentation. Attempting adaptation along with translation of a standardized mathematics achievement test on a priori lines without empirical evidence was not "cost-efficient" in improving test reliability or validity. This does not mean that straight translation was a satisfactory solution, however.

In an attempt to determine the statistical equivalence of test items in a Portuguese translation of a Spanish Test, Clark (1964) made the assumption that equivalence could be demonstrated if corresponding items were symmetric with regard to certain characteristics. Among the characteristics considered relevant were: K-R20 reliability; item difficulty; and item x total test correlation. On the basis of statistical consistencies, Clark concluded that the common linguistic patterns of the two languages permitted the successful item translation/adaptation. He suggested that with less related languages the use of direct translation would be a more difficult undertaking.

Roca (1969) and Renzulli and Paulis (1969) referred to the possible need for reordering of test items in a homogenous test based on an analysis of the correlation between item difficulty and item order. It is the contention of the present authors that if a wide range test such as the PPVT is to be useful, then the basal and ceiling point concept must be operative in the foreign language version. This type of structure presupposes a near perfect correlation between item order item difficulty.

The major purpose of the present study was to determine whether an improved Mexican version of the Peabody Picture Vocabulary Test could be constructed by directly translating both forms A and B of the American test and then using a set of decision procedures to select the better item from the item pair. A re-ordering of the selected items would then be attempted based on an analysis of item difficulties. This resulting single Mexican version would be re-administered to another sample to determine whether or not its psychometric properties were significantly superior to those of the simple translation. A second concern of the study was to test the extent of changes in the psychometric properties of the test when only simple translation was used.

Method

Translating the PPVT

Forms A and B of the PPVT were translated by three bilingual native Mexicans who were members of the faculty of the University of Vera Cruz, Mexico. Each proceeded independently, but in the final stage worked as a group to achieve consensus. These translated forms were referred to as Form A_T and B_T.

First Phase Subjects

Thirty-three students in each of grades one through six were randomly selected from the total school population of Xalapa, Mexico, a metropolitan center of 250,000 inhabitants. Representativeness on sex, school, and morning or afternoon attendance was achieved through

stratification. Translated tests were administered by graduate students in the Department of Psychology of the University of Vera Cruz. Subjects responded to both forms A_T and B_T with a period of one month elapsing between tests. One half of the Ss responded to form A_T first and the other half to B_T to preclude testing effect systematically biasing the results for one of the forms.

Constructing the Revised Form -- PPVT-R_T

Using the ordered criteria listed below, one item was selected from the pair (Forms A_T and B_T) for inclusion in the revised instrument (Form R_T).

1. Point-Biserial Correlation between item score and total test score. This characteristic contributes to the factor structure of the items as well as to test reliability and was regarded as the single most important characteristic of the item. (Nunnally, 1967).
2. When no decision could be made using criterion #1, the point-biserial correlation between item score and chronological age for all subjects in the sample was used as a second necessary characteristic in an attempt to increase test validity.
3. If still no choice was possible, item difficulty as indicated by the mean for each item was examined. In order to maximize reliability, the item of each pair whose pass/fail ratio was closest to .50 was chosen.

The 150 items through the selection process described above were then ordered by difficulty, the mean score for the item. Item difficulties ranged from 1.0 to .10 with a mean of .69 and a SD of .25. The reordering resulted in a pre-cross validation item order by difficulty correlation of .97.

Second Phase Subjects

The revised Peabody, Form PPVT-R_T was administered individually by trained graduate students from the Department of Psychology of the University of Vera Cruz to 120 elementary school Ss from the public schools of Xalapa, Mexico. Twenty subjects at each grade level (one through six) who were not involved in the first phase were randomly selected, with stratification used to control for sex, school, and morning or afternoon attendance. In addition to PPVT-R_T scores, the following data were collected: CA; grade level; school; morning or afternoon session; and, sex.

Results

Alternate form reliabilities were reported in the PPVT test manual (Dunn, 1965). The simple translation of forms A and B resulted in an alternate form reliability of .85 as contrasted with .95 reported for the original test. A statistical test revealed that the reliability of the original form was significantly greater ($z=8.4$; $p .001$).

In table 1 are summarized the results of KR-20 reliability analyses for PPVT forms A_T, B_T, and R_T as well as Z comparisons. Revision of the simple translations based on empirically

Table 1

KR-20 Reliability Estimates and Z
Comparisons for PPVT Form A_T, B_T, and R_T.

Form	KR-20	R _T Comparison Z	p
A _T	.92	1.66	.047
B _T	.92	1.80	.035
R _T	.95		

determined item selection and re-ordering significantly increased the internal consistency of the test.

The correlation between CA and total score was significantly improved by revising the translated versions. Table 2 shows the results of the statistical analyses.

Table 2

Total Score and Chronological Age Pearson Product Moment
Correlation Z Comparison for PPVT
Form A_T, B_T, and R_T.

Form	r	Comparison R _T Z	p
A _T	.42	2.90	.002
B _T	.42	2.85	.022
R _T	.66		

Another comparison concerned the relationship between total PPVT scores and overage for grade. In this analysis, CA was partialled out of the relationship between the dichotomous variable average for grade (1,0) and total test score. Table 3 summarizes the results of this analysis for the simple translations (A_T , B_T) and the revised version (R_T). Again, the revised version evidenced superiority over the simple translation.

Table 3

Partial Correlation Between Total Score and Overage for Grade Classification, Partialling out Chronological Age, Form R_T , A_T and B_T as well as Z Comparisons.

Form	$r_{12.3}$	Z Comparison with Form R_T	p
R_T	-.40		
A_T	-.23	1.73	.018
B_T	-.22	1.83	.033

Discussion

This study was a response to the relative scarcity of psychometrically sound screening devices for Mexican and Mexican-American children. At the same time it was a reaction to the common practice of simply translating a test from one language to another as a means of meeting this need.

The results of this study demonstrate empirically that the reliability of simple translations suffer, if one is willing to assume that these results are typical of similar situations. Since English and Mexican are closely related languages, translation to more distinct languages would be expected to result in even lower reliabilities.

It has been shown that by the use of simple item selection rules and re-ordering, the test developer can economically improve simple test translations when alternate forms of the test are available to form an item pool. Previous work in adapting the Peabody (Brimer and Dunn, 1962) has involved the more expensive, and less cost effective procedure of developing a large number of completely new items as an a priori basis.

Improvement in the correlation between age and total score for the revised test shows that item selection and re-ordering can significantly improve this index of validity, an index whose importance has been mentioned by Dunn (1965) and McNemar (1942) among them. It is worth mentioning that although age by item score correlation was one of the three criteria for item selection, only 38 items were selected on the basis of this information. The fact that the total score by CA correlation was improved was evidence of the usefulness of the procedure.

According to Helmstadter (1964), identification of group membership is one indication of the validity of a test. The partial correlation analysis used in this study represents a very weak source of evidence in this area. Further studies into PPVT-R_T should be addressed to the predictive validity of the test using known mental retardation or giftedness as criteria in a discriminate analysis. Researchers interested in using the PPVT-R_T should contact the authors.

REFERENCES

- Ali, M.B. To establish the feasibility of using translated and adapted versions of an American-made mathematics achievement test in East Pakistan. Dissertation Abstracts, 1967. 28 (1-A), 115-116.
- Clark, John. Item by item parallel achievement test for different foreign languages. Journal of Educational Measurement, Volume 2, Number 1, 1964, 77-83
- Dunn, L.M., Manual for Peabody Picture Vocabulary Test, American Guidance Service, Minnesota, 1965.
- Educational Testing Service. "Test Collection", Princeton, N.J., (1972).
- Helmstadter, G.C. Principles of Psychological Measurement. New York, Appleton-Century-Crofts, 1964.
- Hyness, D. Linguistic aspects of comparative political research, in Holt, R. T. and Turner, J. E. (ed.), op. cit.
- McNemar, Q., The Revision of the Stanford-Binet Scale, an Analysis of the Standardization Data. Boston, Houghton Mifflin, 1942.
- Nunnally, J.C. Psychometric Theory. New York, McGraw-Hill, 1967.
- Psychological Corporation. Test Catalogue. New York, 1972.
- Renzulli, J. S., Paulus, D. H. A cross-validation study of the Item Ordering of the Peabody Picture Vocabulary Test. Journal of Educational Measurement, 1969, 6 (1), 15-20.
- Rieber, M. and Womack, M. The Intelligence of Preschool Children as Related to Ethnic and Demographic Variables. Exceptional Children, 1968, 34, 8, 609-614.
- Roca, P. Problems of adapting intelligence scales from one culture to another. Oficina de Evaluación, Departamento de Instrucción Pública, Halta Rey, Puerto Rico, 1960.