DOCUMENT RESUME

ED 153 386 EC 110 335

AUTHOR Reschly, Daniel J.
TITLE Comparisons of Bias in Assessment with Conventional and Pluralistic Measures.
PUB DATE May 78
NOTE 26p.; Paper presented at the Annual International Convention, The Council for Exceptional Children (56th, Kansas City, Missouri, May 2-5, 1978, Session R5)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cultural Differences; Culture Free Tests; *Definitions; Disadvantaged Youth; Elementary Secondary Education; Exceptional Child Research; *Student Evaluation; *Test Bias; Test Interpretation

ABSTRACT

The paper describes different concepts of bias in tests and presents data on the possible effects of using sociocultural background in the interpretation of standardized test results. Five different definitions of the concept of test bias are explored: equality of means among groups, equal proportions, fairness in predictions, social utility model, and construct validity bias. Assessment data on 1040 children from four ethnic-racial groups (in grades 1, 3, 5, 7, and 9) using conventional and system of multicultural Pluralistic Assessment measures is considered in terms of the above definitions. Among conclusions is that conceptions of test bias are extremely complex and diverse. Tables with statistical data are provided. (SBH)

# [1]Comparisons of Bias in Assessment with Conventional and Pluralistic Measures

Daniel J. Reschly
Department of Psychology
Iowa State University
Ames, Iowa 50011

Litigation in the late 1960's and early 1970's had a significant influence on special education legislation passed by state and federal governments in the mid 1970's. In fact, most of the key requirements of the federal Education for All Handicapped Children Act of 1975 (PL 94-142) can be identified in one or more court decisions over the past ten years. Some of the key PL 94-142 requirements and earlier court decisions are listed below:

Free Appropriate Public Education (PARC, Note 1)

Informed Consent (Diana, Note 2)

Due Process (Diana, Note 2)

Individualized Educational Plan (Guadalupe, Note 3)

Least Restrictive Environment (Guadalupe, Note 3)

Nondiscriminatory Assessment (Diana, Note 2; Larry P., Note 4)

The requirement of nonbiased assessment is one of the most controversial aspects of the recent litigation and legislation. The PL 94-142 Rules and Regulations (Note 5) provide the following statement concerning bias in assessment.

> "Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory." (Section 121a. 530, Part b.)

[1]Presented at the Council for Exceptional Children Annual Convention, May, 1978, Kansas City.

The statement is unequivocal. Racial and cultural discrimination in special education assessment and placement procedures must be eliminated. Unfortunately, the litigation and legislation do not provide clear guidelines concerning the meaning of bias or detailed descriptions of nonbiased assessment and evaluation procedures.

The PL 94-142 Rules and Regulations reflect several implicit assumptions concerning the effects of changes in content and process of assessment and placement procedures. It is assumed that these changes will reduce and perhaps eliminate bias in assessment and placement procedures. Briefly, the changes in content of assessment are: Multifactored assessment in which a broad variety of information is considered including primary language, sociocultural background, and adaptive behavior (Tucker, 1977). The changes in process of assessment and placement include multidisciplinary teams, informed consent, and due process.

The purpose of this paper is to describe different concepts of bias in tests and present data on the possible effects of using sociocultural background in the interpretation of standardized test results.

## Diverse Conceptions of Bias in Tests

The educational and psychological measurement literature contains at least five different definitions of the concept of test bias. These definitions are to varying degrees contradictory and mutually exclusive.

Definition 1. Equality of means among groups. In this definition tests or assessment procedures are defined as biased if different ethnic groups obtain higher or lower scores on the average. The major faults in assessment stressed by these critics have to do with test content or situational factors in assessment (e.g., race of examiner, task demands, etc.) (Jackson, 1975; Williams, 1974). Remedies suggested include development of tests that are more culturally homogeneous, development of pluralistic norms, use of broader varieties of assessment

information (e.g., adaptive behavior outside of school), or in some cases, complete abolition of current standardized tests. Recently, Mercer and Lewis (1978) have developed an approach called SOMPA (System of Multicultural Pluralistic Assessment) which implicitly uses this definition of test bias; the SOMPA provides group specific norms, adaptive behavior information, etc. (Data on the SOMPA for four ethnic-racial groups are presented later.)

Definition 2. Equal Proportions. The second definition requires that the same or nearly the same percentages of persons from different groups be placed in or selected for various programs. That is, if 14% of the population is Native American, then about 14% of the enrollment in EMR (or gifted) programs should be Native American. Overrepresentation of various groups in programs for the mildly retarded has led to litigation. The courts have, at least implicitly, used this definition of test bias in injunctions restraining school districts from placing minorities in programs for the mildly retarded. The remedies required by the courts have included the following: Emphasis on test administration in the child's primary language (Diana and Guadalupe); Lowered cut off scores and use of nonlanguage measures (Guadalupe); And abolition of IQ tests in the diagnosis of mild mental retardation in specific groups (Larry P).

Definition 3. Fairness in Predictions. Two definitions which stress fairness in prediction have been very prominent in the educational and psychological measurement literature (Cleary, 1968; and Thorndike, 1971). Both stress the criterion of equality of prediction, i.e., the same criterion scores are predicted for persons with the same test scores regardless of group membership. However, different methods are used to assess equality of prediction. (Briefly, Cleary suggested analysis of regression equations between groups and Thorndike suggested analysis of number of persons successful on the criterion in relation to number of persons selected by the test.)

Definition 4. Social Utility Model. The fourth definition is relatively recent (Peterson and Novick, 1976), and implies a form of reverse discrimination. The "social utility" of various outcomes would be determined and then test scores would be adjusted in directions that furthered realization of socially desired outcomes. This definition, although provocative, is not directly relevant to special education at this time and hence is not analyzed in the results section of this paper.

Definition 5. Construct Validity Bias. This definition would lead to judgments about test bias on the general criterion of whether the test measures the same traits regardless of group membership. Investigations of factor analysis data, item difficulty indices, and item-score correlations are types of data analyzed in studies of construct validity bias.

## Data on Different Conceptualizations of Test Bias

The data reviewed in this paper were gathered during the Pima County Prevalence Study. Pima County, Arizona is geographically large, ethnically diverse (approximately 68% Anglo, 25% Mexican-American, 4% Black, and 3% Native American) and largely urban (Tucson) with extensive and sparsely populated rural areas.

A stratified random sample of 1040 children was selected with equal numbers from four ethnic-racial groups (Anglo, Black, Chicano, and Native American Papago with N = 260 per group), grade level (1st, 3rd, 5th, 7th, and 9th), sex, and urban-rural residence. A variety of conventional assessment devices were administered to each child in the sample including the Wechsler Intelligence Test for Children - Revised (WISC-R), Metropolitan Achievement Test (MAT), and teacher ratings of classroom achievement and adjustment. In addition to these conventional measures, data were gathered with Mercer and Lewis' System of Multicultural Pluralistic Assessment (SOMPA). Since the SOMPA measures are designed for children between the ages of 5 to 11, SOMPA data were gathered for only three of the five

grade levels in the original sample (grades 1, 3, and 5). A more complete description of the sample and assessment procedures appears in Reschly and Jipson (1976) or Reschly (1978a). The WISC-R, MAT, and Teacher Rating Scales (TRS) were regarded as conventional measures. The SOMPA measures, specifically the Sociocultural Measures (SCM) and Estimated Learning Potential (ELP) scores were regarded as pluralistic measures.

## Results

Definition 1. Equality of means among groups. The nature and magnitude of the differences in mean scores on the WISC-R, MAT, and teacher rating scales among the various groups in the Pima County Prevalence Study closely paralleled differences reported previously in a large number of studies. Reviews and data on these differences are available in a variety of sources (see for example, Sattler, 1974 or Kaufman & Doppelt, 1976). From the perspective of the first definition, all of the conventional measures were biased.

One of the major innovations in SOMPA is the use of pluralistic norms in interpreting the conventional WISC-R results. The pluralistic norms are based on Sociocultural Measures (SCM) which attempt to assess important background variables related to performance on intelligence tests. An individual child's WISC-R score is interpreted in terms of how the child performs in relation to two norm groups. One comparison is based on how the child performed on the WISC-R in relation to the standardization sample. This standard or conventional score is interpreted in SOMPA as the School Functioning Level (SFL). If the child's sociocultural background is significantly different from middle class Anglo patterns, a second score based on pluralistic norms is obtained. The mechanics of obtaining the second score and the underlying rationale are provided by Mercer and Lewis (1978). Briefly, the second score, called Estimated Learning Potential (ELP), is based on adjusting the conventional score through a multiple regression analysis which uses the SCM as predictors.

The SOMPA procedure for computing WISC-R ELP scores eliminates group differences. However, the SOMPA normative data are based on samples of children from California. The authors of SOMPA expressed caution concerning the accuracy of California data for other parts of the country (Mercer and Lewis, 1978). The Pima County Prevalence Study data were analyzed to determine the accuracy of California norms for another geographic area, and to analyze the effects of the ELP score on the first definition of test bias.

In Table 1 the multiple regression equations from the California and Arizona samples for prediction of the WISC-R Full Scale IQ scores are presented.

Although the multiple regression equations in Table 1 appear to be quite different, the ELP scores obtained from the two sets of data are similar (See Table 2). Generally, the ELP scores from other samples will be similar to the California norms <u>if</u> the intercept and multiple correlation of the multiple regression equations are comparable. Data from this study along with data presented by Oakland (1977) on samples of Anglo, Black, and Chicano students in Texas have yielded ELP scores that are relatively close to the California norms.

SOMPA provides the only method known to the author for <u>systematic</u> use of sociocultural background data in special education assessment and placement procedures. Use of sociocultural background data in special education decisions is required in the PL 94-142 Rules and Regulations. Group differences in intelligence test results are either eliminated, or, depending on whether California or local regression equations are used, are greatly reduced by the SOMPA. From the perspective of the first definition of test bias, the SOMPA ELP method of using WISC-R scores is unbiased.

The PL 94-142 Rules and Regulations also require tests and other assessment devices to be <u>valid</u> for the purposes for which they are used. The validity of the SOMPA ELP score is an intriguing, and perhaps in the future, controversial

question. Mercer suggests use of data on acquisition or rate of learning new material as the most appropriate criterion for determining the validity of the ELP score. Conducting studies on acquisition rate is rather difficult and time consuming (Budoff, et al., 1971). Although studies of the relationship of the SOMPA ELP score to conventional measures of achievement are not entirely consistent with the construct of ELP, such studies are useful in clarifying the meaning and appropriate uses of the ELP score in special education decisions.

Two measures of achievement were available from the Pima County Prevalence Study data; Metropolitan Achievement Test Reading and Mathematics subtest scores (MAT-R and MAT-M) and Teacher Rating Scale - Achievement (TRS). Correlations of these conventional achievement measures with the conventional WISC-R (SFL) and pluralistic WISC-R (ELP) scores are presented in Table 3.

A number of interesting trends are apparent in the data presented in Table 3. First the size of the correlations are approximately the same for three of the four groups on both of the types of WISC-R scores. Secondly, the conventional score (SFL) was only slightly better than the pluralistic (ELP) score in predicting achievement. Additional data are needed before firm conclusions are reached, but on the basis of these data it appears that the ELP score may be useful in predicting conventional indices of achievement.

Definition 2. Equal Proportions. The equal selection ratio definition of test bias is very straightforward. It simply requires selection of the same proportions of persons for special programs, etc., that exist in the total population.

The courts have applied this rather simplistic notion of test bias in a number of cases. For example, in the Guadalupe and Larry P. cases (cited earlier) the courts seemed to agree that disproportionate numbers of Non-Anglo students in

EMR programs was a denial of equal protection, and that the ability tests used in the diagnostic process were biased because of the disproportionate ratios. The seemingly simplistic solution of blaming the tests (and indirectly, those who administer them), and in one case, banning the use of such tests, fails to recognize the rather complex process whereby children are referred, evaluated, and sometimes placed in special education programs (See Meyers, Sundstrom, & Yoshida, 1974). Moreover, and most importantly, it fails to deal with the issue of effectiveness of educational programs, whether regular or special, with Non-Anglo students.

With the above cautions in mind, data are presented from a study on the prevalence of mild mental retardation (Reschly and Jipson, 1976). Prevalence of mental retardation was determined from conventional WISC-R results only, i.e., only one dimension of the two dimensional AAMD defnition of mental retardation was used. It should also be noted that the method used to identify the sample, i.e., random selection from school enrollment rosters, is quite different from the referral procedure which is typically the first step in the process through which school age children may be diagnosed as mentally retarded.

These data, although they do not reflect perfectly the real world diagnostic process, provide information on the possible effects of different cut off scores and different IQ scores on the prevalence of mild mental retardation in different groups (See Table 4). The use of the lower cut off score, i.e., 69 rather than 75, led to some reduction in the disproportionality, although the differences in Anglo vs. Non-Anglo percentages were still fairly large. The greatest reduction in disproportionality occurred with the use of the Performance IQ score (P-IQ). Use of P-IQ significantly reduced the disproportionality for Blacks and Native American Papagos, and virtually eliminated it for Mexican-Americans. However, the P-IQ also led to significantly fewer scores below the two cut off scores for Anglos which may be a characteristic of the WISC-R Performance Scale (unlikely) or due to unique characteristics of the present sample.

Pluralistic norms were also used to determine the proportions of children from the four groups who obtained scores below the cut off scores of 70 and 75. Results of using the pluralistic norms, i.e., the SOMPA ELP score, are presented in Table 5. (Note: The data in Table 5 are based on results from grades 1, 3, and 5 only; The n in Table 4 are based on all grades in the sample including grades 7 and 9.)

From the data in Table 5 it appears that use of the SOMPA ELP score as the IQ criterion in decisions about mild mental retardation would result in reducing significantly, but not eliminating potential overrepresentation of minorities in certain special education programs. Use of the ELP score had no effect on the numbers of Anglo children potentially eligible for classification of mild mental retardation. The effects of using the ELP score in this sample were greatest for Native American Papagos, but also significant for Black and Chicano children. Cautions in interpreting these results should again be recognized. The data only represent a small portion of the broad variety of information (including adaptive behavior and primary language) that must be considered in placement decisions. Further, the numbers of children below the respective cut off scores in this study were rather samll in some cases.

Definition 3. Equality of Prediction. To date we have conducted one study which provides data on the potential bias of the WISC-R and MAT in terms of the third definition (Reschly and Sabers, 1978). In this study the Cleary definition of test bias was used to examine the equivalence of predictions across the groups using the MAT and WISC-R as the criterion and predictor respectively. These comparisons involved examination equality of errors of estimate, slopes, and intercepts of regression equations for each group at the five grade levels.

Generally, the regression equations for the different groups were unequal with the majority of the differences arising from unequal intercepts or unequal

slopes. In cases in which the prediction systems differed due to unequal errors of estimate, the errors of estimate were consistently smaller for the Non-Anglo groups. The direction of differences in slope were about equally divided with the Anglo slope being higher for some comparisons, but lower for others. The clearest differences, and perhaps the most significant in practical terms, were the differences in intercept. Differences in intercept lead to different predicted scores for individuals who in fact have the same scores on the predictor measure (Anastasi, 1976). These differences are clearly sufficient to establish the existence of bias in a technical sense, and may also be "unfair" in a practical sense if different groups gain differential access (or vulnerability) to positive or negative circumstances on the basis of the prediction.

Intercept differences lead to over or under prediction for at least some of the groups. The direction of over and under prediction for these groups using a common regression line was analyzed (See Table 6). In nearly all cases the outcome of the common regression line was over prediction for Non-Anglo groups and under prediction for Anglos, a result which is consistent with previous literature (Stanley, 1971). The amount of over prediction for the Non-Anglo groups would have been even greater if the regression equation based on Anglos only would have been applied to all groups.

Definition 4. Social Utility. We have conducted no studies which would provide data on the fourth definition of bias in tests. Data on the fourth definition is extremely difficult if not impossible to generate. The social utility of placement in various educational alternatives has yet to be determined, and is likely to be controversial as such. For example, is it beneficial or harmful to be determined eligible for and to receive services under such classifications as learning disability? remedial reading? educationally handicapped?, etc.

Definition 5. Construct Validity Bias. A large number of possible types of studies could be generated which would provide data on the construct validity of various tests among various groups. A comparison of the WISC-R factor structures among the groups in the Pima County Study is the only analysis of this type that we have conducted to date (Reschly, 1978a). Generally, the WISC-R factor structures were highly similar in the two factor solution, but dissimilar for a three factor solution. The objective evidence on the number factors that "should" be identified on the WISC-R for the various groups was inconsistent. The available evidence supported a three factor solution for Anglos and Chicanos, and two factor solutions for Blacks and Native American Papagos (See Tables 7 and 8).

If we use the two factor solutions, and interpret only these factors on the WISC-R, then the conclusion of no bias would be supported by the data. The picture for the three factor solutions is more complex. The first two factors in the three factor solution were highly similar across the groups. The third factor was different, especially for Blacks and Native American Papagos. For these and other reasons (Reschly & Reschly, in press) caution should be used in any interpretation of the third WISC-R factor.

## Discussion

The data provided in this paper can obviously be used to support a variety of divergent conclusions regarding the overall question of test bias. The clearest and most important conclusion is that conceptions of test bias are extremely complex and diverse. Due to the complexity and diversity of conceptions of test bias, unequivocal or simple yes-no answers to the question of test bias are impossible.

The kind of definition of test bias used is a clear influence on the outcome of any analysis of test bias. In this paper five different conceptions of test bias have been discussed. These definitions were used in varying degrees

as the basis for analyses of data, and not surprisingly, led to different conclusions regarding test bias. It is important for us to recognize the influence of how test bias is defined, and to formulate clearly the definition of test bias used in future discussions of this issue.

Secondly, we must recognize that test bias, even if clearly defined, will always be a matter of degree and dependent on situational conditions. Just as tests are never "valid" in any global or all encompassing sense, a specific test is not simply biased or unbiased. The bias is always a matter of degree and further dependent on such variables as age, group, setting, purpose, etc.

Mercer and Lewis (1978) contend that the SOMPA procedures will reduce bias in assessment procedures. The data presented in this paper support the conclusions that SOMPA is less biased in terms of the first two definitions of test bias (equal means and equal proportions). Reductions in number of students, especially minority, eligible or classified for special education is one of the possible outcomes of widespread adoption of SOMPA. Declassification in and of itself may not be particularly beneficial to children. SOMPA is extremely complex. The system involves much more than simply adjusting scores for culturally different children. The ultimate potential of changes in assessment and placement procedures such as SOMPA and the multifactored assessment model (Tucker, 1977) is more refined classification and intervention procedures. Some examples may illustrate this point. Use of the SOMPA pluralistic norms may lead to less segregation of minority students. Use of the adaptive behavior data may lead to selection of more appropriate service options (Reschly, 1978b) and identification of appropriate goals for changes in social behaviors. Use of the multifactored assessment information may be a tool for development and then selection of a variety of service options (Deno, 1972). The challenge before us is not elimi-

nating assessment procedures, but developing more refined and precise methods of gathering information, and then translating this information into effective interventions.

The debate on nonbiased assessment has led to the development of conceptions of bias that are broader than narrow concerns with specific tests. Ysseldyke (in press) describes the kinds of bias that arise from naturally occurring characteristics of students such as attractiveness, socioeconomic status, etc. These forms of bias occur apart from or after formal assessment procedures. Recognition of these sources of bias is a prerequisite to effective procedures for insuring fairness in special education assessment and placement procedures.

The most important step in eliminating bias must be effective educational interventions. In the view of the present author, an outcomes criterion must guide our overall effort to achieve fairness in special education assessment and placement procedures (Reschly, 1978b). Improved assessment practices and more refined educational alternatives are prerequisites to the goal of greater effectiveness. The use of pluralistic assessment procedures as a supplement to conventional assessment practices appears to have considerable potential for moving us toward more precise and effective interventions for children.

Table 1

Comparison of Arizona and California Multiple Regression Equations for Predicting WISC-R Full Scale IQ

| | | |
|---|---|---|
| ANGLO | CA WISC-R FS IQ = 79.77 + 1.5 SES - .42FSI + .14 UA + .32 FST | Multiple R = .42 |
| | AZ WISC-R FS IQ = 82.05 + .65 SES - .19 FSI + .23 UA - .04 FST | Multiple R = .30 |
| BLACK | CA FS IQ = 76.83 + .49 SES - .46 FSI + .19 UA + .22 FST | Multiple R = .37 |
| | AZ FS IQ = 75.97 + .69 SES - .27 FSI + .11 UA + .23 FST | Multiple R = .34 |
| CHICANO | CA FS IQ = 84.86 + .42 SES - .29 FSI + .20 UA + 0.0 FST | Multiple R = .39 |
| | AZ FS IQ = 83.13 + .84 SES - .54 FSI + .11 UA + .01 FST | Multiple R = .36 |
| NATIVE AMERICAN PAPAGO | AZ FS IQ = 61.91 - .68 SES + .19 FSI + .32 UA + .13 FST | Multiple R = .32 |

CA = California data from Mercer and Lewis, 1978.
AZ = Arizona data from Pima County Prevalence Study
SES = Socioeconomic Status Score From SOMPA Sociocultural Scales
FSI = Family Size Score from SOMPA Sociocultural Scales
UA = Urban Acculturation Score From SOMPA Sociocultural Scales
FST = Family Structure Score From SOMPA Sociocultural Scales

Table 2
Comparison of Estimated Learning Potential Scores Derived From Arizona and California Multiple Regression Equations

| GROUP | WISC-R Score | ELP Mean AZ Formula | ELP Mean CA Formula | ELP S.d. AZ Formula | ELP S.d. CA Formula | ELP Range AZ Formula | ELP Range CA Formula | MEAN of Differences AZ ELP-CA ELP | Mean[1] Difference AZ ELP-AZ SFL | Range[1] AZ ELP-AZ SFL |
|---|---|---|---|---|---|---|---|---|---|---|
| | Verbal | 101.35 | 100.76 | 15.19 | 15.68 | 46-149 | 46-149 | 0.59 | 2.12 | 0 - 15 |
| ANGLO | Performance | 101.55 | 100.82 | 13.75 | 13.13 | 52-140 | 55-133 | 0.73 | 0.93 | 0 - 11 |
| | Full Scale | 101.54 | 100.70 | 14.23 | 13.48 | 52-140 | 55-140 | 0.84 | 1.68 | 0 - 16 |
| | Verbal | 100.05 | 96.77 | 15.75 | 16.00 | 56-146 | 51-146 | 3.28 | 14.27 | 0 - 29 |
| BLACK | Performance | 100.30 | 99.12 | 15.27 | 13.55 | 53-136 | 58-133 | 1.18 | 11.41 | 0 - 25 |
| | Full Scale | 100.20 | 96.74 | 15.75 | 14.73 | 56-147 | 56-142 | 3.46 | 14.09 | 0 - 29 |
| | Verbal | 100.03 | 94.22 | 15.98 | 13.74 | 61-138 | 60-127 | 5.81 | 15.42 | 0 - 29 |
| CHICANO | Performance | 99.82 | 93.66 | 15.36 | 13.98 | 62-137 | 59-127 | 6.15 | 7.52 | 0 - 18 |
| | Full Scale | 99.91 | 92.93 | 15.83 | 13.89 | 60-136 | 62-122 | 6.98 | 12.79 | 0 - 26 |
| NATIVE | Verbal | 99.83 | Not | 15.59 | Not | 59-147 | Not | Not | 25.53 | 7 - 41 |
| AMERICAN | Performance | 99.84 | Avail- | 15.47 | Avail- | 65-143 | Avail- | Avail- | 13.42 | 0 - 28 |
| PAPAGO | Full Scale | 100.11 | able | 15.55 | able | 56-136 | able | able | 21.39 | 1 - 37 |

All data in this table are based on the Pima County Prevalence Study, Grades 1, 3, & 5.

[1]The last two columns provide information on the average amount of difference between the pluralistic (ELP) and conventional (SFL) WISC-R scores, and the range of the differences between ELP and SFL.

Table 3

Relationship of SOMPA SFL and ELP Scores to Conventional Measures of Achievement

| | | MAT-R | MAT-M | TRS-ACH |
|---|---|---|---|---|
| ANGLO | SFL-Verbal (V) | .50 | .49 | .37 |
| | ELP-V | .49 | .49 | .34 |
| | SFL-Performance (P) | .36 | .33 | .22 |
| | ELP-P | .36 | .33 | .22 |
| | SFL-Full Scale (FS) | .51 | .48 | .34 |
| | ELP-FS | .51 | .49 | .33 |
| BLACK | SFL-V | .66 | .54 | .51 |
| | ELP-V | .66 | .55 | .49 |
| | SFL-P | .37 | .48 | .28 |
| | ELP-P | .35 | .47 | .25 |
| | SFL-FS | .61 | .58 | .46 |
| | ELP-FS | .61 | .59 | .44 |
| CHICANO | SFL-V | .53 | .38 | .39 |
| | ELP-V | .45 | .34 | .35 |
| | SFL-P | .45 | .35 | .45 |
| | ELP-P | .40 | .29 | .39 |
| | SFL-FS | .57 | .42 | .47 |
| | ELP-FS | .49 | .36 | .43 |
| NATIVE AMERICAN PAPAGO | SFL-V | .33 | .41 | .41 |
| | ELP-V | .29 | .36 | .35 |
| | SFL-P | .33 | .39 | .29 |
| | ELP-P | .33 | .35 | .26 |
| | SFL-FS | .37 | .45 | .40 |
| | ELP-FS | .35 | .40 | .35 |
| ALL GROUPS COMBINED | SFL-V | .65 | .60 | .42 |
| | ELP-V | .43 | .40 | .39 |
| | SFL-P | .50 | .49 | .35 |
| | ELP-P | .33 | .33 | .29 |
| | SFL-FS | .65 | .61 | .43 |
| | ELP-FS | .44 | .41 | .39 |

Table 4

[2]Group Differences in Proportions of WISC-R IQ Scores
Below Cut Off Scores of 69 and 75

| IQ Score/Group | IQ ≤ 69 % below cut off | IQ ≤ 69 Disproportionality[1] | IQ ≤ 75 % below cut off | IQ ≤ 75 Disproportionality[1] |
|---|---|---|---|---|
| Verbal IQ | | | | |
| Anglo | 2.4 | | 4.8 | |
| Black | 10.2 | 4.25:1 | 22.1 | 4.60:1 |
| Mexican-American | 10.8 | 4.50:1 | 24.2 | 5.04:1 |
| Native American Papago | 37.5 | 15.63:1 | 60.8 | 12.67:1 |
| Performance IQ | | | | |
| Anglo | 1.2 | | 2.0 | |
| Black | 4.7 | 3.92:1 | 12.3 | 6.15:1 |
| Mexican-American | 2.2 | 1.83:1 | 8.9 | 4.45:1 |
| Native American Papago | 4.2 | 3.50:1 | 15.8 | 7.90:1 |
| Full Scale IQ | | | | |
| Anglo | 1.6 | | 2.4 | |
| Black | 8.1 | 5.06:1 | 16.6 | 6.92:1 |
| Mexican American | 6.7 | 4.19:1 | 16.1 | 6.71:1 |
| Native American Papago | 14.2 | 8.88:1 | 37.1 | 15.46:1 |

[1]Where disproportionality is computed by dividing NonAnglo Percentage by the Anglo Percentage.

[2]Author's Note: The percentages contained in this table should not be viewed as indicative of the "real" prevalence of actual mental retardation among these groups. At most, the percentages reflect what might be called "psychometric mental retardation." For further discussion of this distinction, see Reschly and Jipson (1976), Grossman (1973), or Mercer (1973).

Results based on grades 1, 3, 5, 7, and 9.

Table 5

[1]Perceptage of WISC-R-Full Scale Scores Below
Cut Off Scores of 70 and 75 on SOMPA SFL and ELP

| | Cut Off Score of ≤ 70 | | Cut Off Score of ≤ 75 | |
|---|---|---|---|---|
| | SFL-FS | ELP-FS | SFL-FS | ELP-FS |
| ANGLO N = 149 | 2.7% N = 4 | 2.7% N = 4 | 3.4% N = 5 | 3.4% N = 5 |
| BLACK N = 128 | 7.8% N = 10 | 4.7% N = 6 | 13.3% N = 17 | 8.6% N = 11 |
| CHICANO N = 125 | 9.6% N = 12 | 2.4% N = 3 | 20.8% N = 26 | 6.4% N = 8 |
| NATIVE AMERICAN PAPAGO N = 122 | 22.9% N = 28 | 3.3% N = 4 | 36.9% N = 45 | 6.5% N = 8 |

[1]Based on data from the Pima County Prevalence Study, Grades 1, 3, and 5

Table 6

Actual Mean and Predicted Mean Achievement Scores
For Four Ethnic-Racial Groups Based on a Common Regression Equation
Using WISC-R Full Scale and MAT Reading and Mathematics

| Grade | Group[1] | N | Predicted Reading (P) | Actual Reading (A) | $\frac{P-A}{s.d.}$[2] | Predicted Math (P) | Actual Math (A) | $\frac{P-A}{s.d.}$[2] |
|---|---|---|---|---|---|---|---|---|
| 1st | Anglo | 49 | 54.09 | 56.06 | -.21 | 54.37 | 56.29 | -.22 |
| | Black | 40 | 50.14 | 48.52 | .19 | 50.16 | 47.48 | .30 |
| | C | 44 | 50.57 | 52.95 | -.24 | 50.61 | 52.45 | -.19 |
| | NAP | 48 | 46.87 | 44.04 | .38 | 46.67 | 45.27 | .17 |
| 3rd | Anglo | 51 | 55.16 | 55.97 | -.07 | 54.02 | 55.68 | -.18 |
| | Black | 40 | 49.10 | 52.20 | -.37 | 49.18 | 49.41 | -.03 |
| | C | 45 | 49.51 | 50.28 | -.09 | 49.51 | 51.37 | -.19 |
| | NAP | 51 | 45.57 | 41.66 | .71 | 46.36 | 42.88 | .49 |
| 5th | Anglo | 52 | 56.99 | 58.40 | -.13 | 56.74 | 57.16 | -.04 |
| | Black | 45 | 48.17 | 47.48 | .08 | 48.22 | 48.56 | -.04 |
| | C | 48 | 49.54 | 48.59 | .11 | 49.53 | 50.91 | -.17 |
| | NAP | 44 | 44.84 | 44.92 | -.01 | 44.99 | 42.64 | .38 |
| 7th | Anglo | 54 | 55.99 | 57.48 | -.17 | 55.40 | 55.92 | -.05 |
| | Black | 51 | 49.15 | 48.95 | .02 | 49.19 | 50.29 | -.12 |
| | C | 46 | 49.61 | 48.96 | .08 | 49.62 | 50.12 | -.06 |
| | NAP | 43 | 44.19 | 43.24 | .13 | 44.70 | 42.20 | .47 |
| 9th | Anglo | 44 | 57.38 | 59.32 | -.22 | 57.05 | 59.51 | -.23 |
| | Black | 46 | 47.74 | 47.35 | .04 | 47.86 | 46.73 | .18 |
| | C | 32 | 49.33 | 50.02 | -.08 | 49.37 | 49.87 | -.05 |
| | NAP | 37 | 46.13 | 43.70 | .36 | 46.32 | 44.37 | .37 |

[1] C refers to Chicano and NAP refers to Native American Papago.

[2] Predicted mean less the actual mean divided by the obtained subgroup standard deviation.

Table 7

WISC-R Subtest Loadings in Three Factor Solution for Four Ethnic Groups

| Group | ANGLO | | | BLACK | | | CHICANO | | | NATIVE AMERICAN PAPAGO | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor | I | II | III | I | II | III | I | II | III | I | II | III |
| WISC-R Subtest | | | | | | | | | | | | |
| I | 63 | 32 | 26 | 66 | 40 | 18 | 66 | 20 | 33 | 68 | 22 | 21 |
| S | 59 | 26 | 26 | 59 | 41 | 13 | 67 | 15 | 22 | 58 | 33 | 11 |
| A | 43 | 26 | 45 | 61 | 34 | 27 | 40 | 13 | 45 | 42 | 37 | 09 |
| V | 74 | 23 | 12 | 75 | 20 | 16 | 67 | 26 | 30 | 74 | 15 | 05 |
| C | 64 | 22 | 21 | 71 | 24 | 09 | 61 | 20 | 06 | 70 | 10 | 17 |
| DS | 35 | 02 | 40 | 49 | 08 | 36 | 33 | 14 | 31 | 30 | 35 | 09 |
| PC | 20 | 49 | 09 | 25 | 52 | 21 | 32 | 52 | 12 | 21 | 53 | 14 |
| PA | 20 | 53 | 00 | 29 | 53 | 24 | 17 | 38 | 39 | 23 | 44 | 03 |
| BD | 17 | 60 | 22 | 20 | 33 | 58 | 20 | 59 | 16 | 14 | 69 | 05 |
| OA | 07 | 59 | 18 | 10 | 17 | 58 | 14 | 58 | 09 | 07 | 51 | 25 |
| Co | 12 | 16 | 40 | 33 | 20 | 22 | 14 | 16 | 37 | 17 | 17 | 37 |
| M | 18 | 42 | 10 | 23 | 44 | 30 | 06 | 47 | 20 | 14 | 51 | 28 |

Note. All decimal points have been omitted.

Table 8

Coefficients of Congruence for Two and Three

Factor Solutions for Four Groups

Two Factor Solutions

| | Black | | Chicano | | Native American Papago | |
|---|---|---|---|---|---|---|
| | Factor I | Factor II | Factor I | Factor II | Factor I | Factor II |
| Anglo | .99 | .97 | .99 | .98 | .99 | .97 |
| Black | | | .99 | .98 | .98 | .99 |
| Chicano | | | | | .98 | .97 |

Three Factor Solutions

| | Black | | | Chicano | | | Native American Papago | | | [1]Standardization Data | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | I | II | III | I | II | III | I | II | III |
| Anglo | .98 | .91 | .76 | .99 | .98 | .86 | .99 | .95 | .78 | .98 | .98 | .97 |
| Black | | | | .97 | .89 | .74 | .98 | .89 | .73 | .96 | .89 | .76 |
| Chicano | | | | | | | .99 | .96 | .72 | .98 | .99 | .93 |
| Native American Papago | | | | | | | | | | .98 | .96 | .74 |

[1]Coefficients reported are based on comparison o~ loadings from this study with the median loadings for the varimax rotation reported by Kaufman (1975, Table 4, p. 141).

## Reference Notes


1. Pennsylvania Association for Retarded Children, Nancy Beth Bowman et al. v. Commonwealth of Pennsylvania, David H. Kurtzman et al. Civil Action No. 71-42 (3 Judge Court, E. D. Pennsylvania, 1971).
2. Diana v. State Board of Education, C-70 37 RFP, District Court for Northern California (February, 1970).
3. Guadalupe v. Tempe Elementary School District, 71-435, District Court for Arizona, January, 1972.
4. Larry P. et al. v. Wilson Riles et al. United States District Court, Northern District of California, Case No. C-71-2270 RFP.
5. Rules and Regulations for Public Law 94-142, Federal Register, August 23, 1977.

## References


Anastasi, A. Psychological testing. (4th ed.) New York: Macmillan, 1976.

Budoff, M., Meskin, J., & Harrison, R. Educational test of the learning-potential hypothesis. American Journal of Mental Deficiency, 1971, 76, 159-169.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Deno, E. Special education as developmental capital. Exceptional Children, 1972, 37, 229-237.

Jackson, G. On the report of the ad hoc committee on educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 88-92.

Kaufman, A. Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. Journal of Consulting and Clinical Psychology, 1975, 43, 135-147.

Kaufman, A. & Doppelt, J. Analysis of WISC-R standarization data in terms of stratification variables. Child Development, 1976, 47, 165-171.

Mercer, J. & Lewis, J. Technical Manual: SOMPA. New York: Psychological Corporation, 1978.

Meyers, C., Sundstrom, P., & Yoshida, R. The school psychologist and assessment in special education: A report of the Ad Hoc Committee of APA Division 16. Monographs of Division 16 of the American Psychological Association, 1974, 2(1), 3-57.

Oakland, T. Pluralistic norms and estimated learning potential. Annual Convention of the American Psychological Association, August, 1977.

Peterson, N. & Novick, M. An evaluation of some models for culture fair selection. Journal of Educational Measurement, 1976, 13, 3-29.

Reschly, D. WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 1978a, in press.

Reschly, D. Nonbiased assessment and school psychology. Des Moines, IA: Iowa Department of Public Instruction, 1978b. (Mimeo, 96 pgs.)

Reschly, D. & Jipson, F. Ethnicity, geographic locale, age, sex, and urban-rural residence as variables in the prevalence of mild retardation. American Journal of Mental Deficiency, 1976, 81, 154-161.

Reschly, D. & Reschly, J. Predictive utility of WISC-R factor scores for four groups. Journal of School Psychology, in press.

Reschly, D. & Sabers, D. An examination of test bias for Black, Chicano, and Native American Papago Children. Unpublished manuscript, 1978.

Sattler, J. Assessment of children's intelligence. Philadelphia: W. B. Saunders, 1974.

Stanley, J. Predicting college success of the educationally disadvantaged. Science, 1971, 171, 640-647.

Thorndike, R. Concepts of culture fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Tucker, J. Operationalizing the diagnostic-intervention process. In T. Oakland (Ed.) Psychological and Educational Assessment of Minority Children. New York: Brunner/Mazel, 1977.

Williams, R. The problem of the match and mismatch in testing black children. In L. Miller (Ed.), The Testing of Black Students: A Symposium. Englewood Cliffs, NJ: Prentice-Hall, 1974.

Ysseldyke, J. Issues in psychoeducational assessment. In Phye, G. & Reschly, D. School Psychology: Perspectives and issues. New York: Academic Press, in press.