

The Standardization of Intelligence Tests On Southern Negro School Children

Frank Garfunkel and Burton Blatt

ABSTRACT

The task of predicting the academic success of Southern Negro school children is complicated because the school and its relationship to vocational accomplishment represent, in general, a different set of values for Southern Negroes than for other cultural groups. Psychometric instruments, such as the Stanford-Binet Test, cannot necessarily be indiscriminantly used with all cultural groups. A recent study that makes a straightforward application of the Stanford-Binet Test to develop a set of norms for Southern Negro school children is criticized in that it assumes that the test is appropriate and all that is needed is the development of norms. The criteria for success and the means for attaining it must be reevaluated at the same time that predictive instruments are developed.

A recent monograph (Kennedy, Van De Riet and White, 1961) describes the results of the administration of the Stanford-Binet Intelligence Test (1960) and the California Achievement Test Battery (1951) to 1800 Negro elementary-school children in five Southeastern states: Florida, Georgia, Alabama, Tennessee, and South Carolina. The children were randomly selected from representative counties which were stratified by county size and by school grade. In addition to the test results, data was obtained on teacher ratings of academic performance and on various demographic variables. The critical importance of this study revolves around two pressing problems: (1) The exploration of intellectual variation in underprivileged socio-cultural groups, and (2) the application of standardized testing instruments to groups strikingly different from those upon which standardization was accomplished. The question to be raised in this paper is whether either of these problems has been adequately dealt with in a study which has been designed ". . . to provide educators, counselors, clinicians and researchers with the normative data essential for their work with these school children" (Kennedy, et al., 1961).

Kennedy and his associates have insightfully summarized those problems that have interfered with the interpretation and evaluation of intelligence test scores of Negroes, using presently available instruments such as the 1960 revision of the Stanford-Binet. Their convictions that "race" isn't of particular significance in contributing to the variability of intelligence test scores and that differences in test performance result, to a great extent, from variables in the environment as well as from inherited or constitutional differences are, as they have indicated, well documented in the literature. Furthermore, their discussion of the problems of inadequate standardization, the implications of both socio-economic and caste status,

FRANK GARFUNKEL, Ed.D., is professor, Education Department, University of Connecticut, Storrs, Conn. BURTON BLATT, Ed.D., is Professor and Chairman, Special Education Department, Boston University, Boston, Mass.

the probable language variation in relatively isolated groups, differential educational opportunity, motivation to take tests and selective migration, would appear to be in accord with the literature that they review (1961, p. 34-35).

The Problem of constructing norms for sub-cultural groups¹

It is the position of this paper that while the results of this study provide a careful confirmation of what is relatively well known about the tested intelligence and achievement of children from lower social classes, it promulgates a rather serious misconception about the relationship between a global measure of intelligence and social class when the former has not been standardized on the latter. The misconception revolves around the proposition that "validity cannot be regarded as a fixed or a unitary characteristic of a test" (Gullikson, 1950, p. 88). It depends to a considerable extent, on a specific criterion. Such a criterion has been used in the development of the Stanford-Binet Intelligence Test but it does not necessarily apply to groups other than those on whom the test was standardized. Because a test is valid for one group does not mean that it will be valid for other groups, no matter how carefully norms are established and no matter how reliable the test is reputed to be. The misconception is the assertion that the Stanford-Binet Intelligence Test (or any other intelligence test) is necessarily a valid measure of the intelligence of children from lower social classes and, in particular, of Southern Negro Children. This problem is not dealt with in the study being reviewed. Furthermore, there is every reason to believe that the culture of Southern Negroes is in such contrast to that of middle class whites that the test under consideration would probably not be valid as a prediction of academic success. Kennedy's statement that ". . . when one cultural group is administered an intelligence test which has been constructed for, and standardized on another cultural group, the former consistently scores lower" (p. 35), does not necessarily imply that there is a transformation possible between the norms of the standardization group and the individuals of the other cultural group. Although such a consistent difference in means is a necessary condition, it is not a sufficient one. In order to establish the validity of a transformation, as represented by conversion tables, it would be necessary to also show consistently correlated individual variation — that the inter-group variation is related to inter-individual variation.

¹ The term "sub-cultural" is used here to describe those groups that present such socio-cultural characteristics as to create problems related both to their adjustment and integration into the "major" culture and the abilities of that "major" culture to tolerate them in school, in the community, and in other instances of human experience. The selection of this term may, in view of its derogating tone, be objectionable to some. It is here used in a desire to facilitate communication. We view the term "multicultural" as more acceptable connotatively and more descriptive sociologically.

The problem is not simply one of not including a particular minority group in a standardization sample. If this was so then the procedures of this study would probably be appropriate. But scores on intelligence tests (or achievement tests) are not precise measurements of specific behavior that is independent of environmental (standardization) variables, and whose base lines can be determined by testing additional individuals or groups with a given instrument. The development of these tests is dependent upon a complex matrix of base lines including previous learning experiences, test taking abilities, and success criteria. For example, items on the Stanford-Binet have been selected in accordance with pass-fail criteria for different age groups from certain environments. Selected items are those that best discriminate between more and less academically successful children who are in certain kinds of schools and homes and, consequently, have different kinds of successful experiences. It is hardly justifiable to simply give these children a test that has been developed in another milieu and expect the results to be at all comparable. As a matter of fact, there is rather convincing evidence that the results are not comparable. This evidence is best illustrated in follow-up studies that have been done on mildly mentally retarded children after they leave school. They appear to have adjusted as well and to be as successful as their social class peers in spite of the fact that their tested intelligence was markedly inferior (Clarke and Clarke, 1958, Chapter 17).

Post-school success is not necessarily synonymous with adequate school performance and, *in the particular groups under consideration*, it is fairly safe to say that successful performance in life is independent of school success. Conversion tables can do much to obscure this logical and well documented condition. Although Kennedy and his co-authors allude to the problem concerning just what is accomplished with such "special" norms, one has the vague unsatisfied notion that norms of this type may confuse rather than clarify an understanding of the intellectual potentials of Negro children. Although it would be salutary to develop meaningful and convertible norms for different groups, it is a questionable task when the experiences of these groups are so different as to suggest remarkably different educational and vocational goals.

Another relevant question is suggested by the authors' statement that ". . . a Negro child who, in spite of the cultural deficits facing him, scores an I. Q. of 100 must be a superior child indeed to surmount these enormous difficulties" (p. 143). The purpose of intelligence tests is not to provide a regression equation for socio-cultural deprivation but to obtain a useful prediction for children upon whom it was standardized. The clinical judgment suggested by this quote does violence to several incisive questions. What difficulties in motivation did this particular child surmount? How is this child different from white children with

I. Q.'s of 100 or Negro children with higher or lower I. Q. scores? Is the Negro child with an I. Q. of 80 (the approximate mean for the total sample) equivalent in intellectual potential to a white child with an I. Q. of 100? What is the relation between school and non-school problem solving tasks and what is the effect of this relationship on intellectual performance?

Methodological Problems

The utilization of the tables of this study (Kennedy, et al, 1961) is further complicated by several restrictions that evolve from the sampling design employed for the selection of subjects. The regular standardization of the Stanford-Binet (1960) was made by selecting children from chronological age strata and not from school grade strata. The Negro standardization has been made by selecting Negro elementary-school children from six school-grade strata and includes ages five through sixteen. This inappropriate deviation in sampling procedure becomes especially pertinent when analyzing data concerning the youngest and oldest children in the study. For example, although 86.7% of the thirteen year olds in the study had I. Q.'s of 75 or less, this says little or nothing about the population of thirteen year old Negro children, i. e. those who are not in the fifth or sixth grade, but who are in grades seven, eight and nine. This criticism can be applied to the twelve year olds (57.8% had I. Q.'s of 75 or less) and to a lesser extent, to the eleven year olds (36.6% had I. Q.'s of 75 or less). At the other end of the age scale the same phenomena occurs, but in reverse. This study does not sample the five year old population, but only those in school. The more sub-normal five year olds may not be in school, and, therefore, the 5.3% with I. Q.'s of 75 or less is clearly an underestimate.

This obvious discrepancy between the regular standardization and this special one points to the speculation that inferences made from the data on the intermediary grade strata are subject to a bias that is obscured by the oversimplified analysis that was done. In the analysis of the data by age and grade on one hand, and by socio-economic level and community size on the other, there was no attempt to describe effects other than single dimensional variability of each of the variables under consideration. This is unfortunate because the study of these interactions would be vital to an understanding of what is happening psychometrically to children in substandard schools. Since this is a question that is not asked by the authors, it is not surprising that they don't look at their data in such a way as to suggest explanations. Perhaps, the reason for this oversight is inherent in the fallacy of their approach of trying to convert data on noncomparable groups.

Discussion

Even with the restrictions discussed in the above two sections, there are certainly meaningful data that can provide some insights into educational conditions in Southern Negro elementary schools. With regards to the six to eleven year-olds, the figures are consistent with what has been found in urban lower class school districts. For example, Fouracre (1961, p. 46-47) found that 20% to 30% of school age children in lower class school districts of Pittsburgh had I. Q.'s in the mentally retarded range. This compares to 20% to 40% in the Kennedy, et al, study (1961). Another way of stating this is in terms of yearly increments of Stanford-Binet mental ages. For those ages that are probably fairly well represented in the sample, six to eleven years, the mean yearly increments (in months) are 8, 9, 9, 10, and 7, respectively, with an overall mean of 8.6 months. This compares with the 12 month increment that exists for the regular standardization sample. The yearly mental age increments of the school grade levels are more or less meaningless because of the impossibility of ascertaining school placement procedures.

Conclusion

In the light of the above, it should be fairly apparent that the standardization under discussion raises many challenging questions about the measurement of the intellectual abilities of children from lower social classes. There is serious question as to whether this particular study does more than confirm what is fairly well documented in the literature—that children from lower social classes test as subnormal levels on standardization measures of intellectual performance. Because of the restrictions of the sampling design, it is doubtful that the conversion tables constructed on the basis of the testing of 1800 Negro elementary-school children are practically meaningful. In spite of the authors' assertion that "No attempt will be made to answer the questions, 'What is intelligence?' or 'What aspects of intelligence are innate and what are learned?'" (p. 3) there are serious misgivings as to whether they have or whether they could possibly treat the "practical" questions that they list as their goals, without giving these theoretical problems careful consideration. What they have done is collect a great deal of data in a more or less convenient way and presented it in a straightforward manner without any clear regard for the Stanford-Binet standardization or for the limitations of their approach. It is not enough to simply state that "This is all that was done, this and no more." Studies such as this should be based on an appropriate theoretical model that is consistent with underlying psychological principles.

References

- Clark, A. M. and Clarke, A. D. B. *Mental Deficiency: The Changing Outlook*. Glencoe, Illinois: Free Press, 1958.
- Fouracre, M. H., Rooke, M. L. and Botwin, P. *The Report of the Study on the Educational Needs of Physically Handicapped Children in Pittsburgh, Pennsylvania, 1958-59*. Pittsburgh: University of Pittsburgh, 1961.
- Kennedy, W. A., Van De Riet, V. and White, J. C. *The Standardization of the 1960 Revision of the Stanford-Binet Intelligence Scale on Negro Elementary-School Children in the Southeastern United States*. Tallahassee, Florida: Florida State University, 1961.
- Terman, L. M. and Merrill, M. A. *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin, 1960.
- Tieg, E. W. and Clark, W. W. *California Achievement Tests*. Los Angeles: California Test Bureau, 1951.

