AN UNFOUNDED CONCLUSION IN M. W. SMITH'S ANALYSIS OF CULTURE BIAS IN THE STANFORD-BINET INTELLIGENCE SCALE* [1]

University of California, Berkeley

ARTHUR R. JENSEN

SUMMARY

This article is a critique on M. W. Smith's "Alfred Binet's remarkable questions: A cross-national and cross-temporal analysis of the cultural biases built into the Stanford-Binet Intelligence Scale and other Binet tests," published in *Genetic Psychology Monographs*, 1974, **89**, 307-334.

A. INTRODUCTION

A main conclusion of an article in a recent issue of *Genetic Psychology Monographs* is based on serious factual errors and is therefore wholly unfounded. Smith (2) concludes that when another population is administered the Binet test, its scores are lower than those of the original normative sample and that *this is true also for a group with the same characteristics of the normative sample but tested at a later point in time*. Smith claims that the lowering of test scores over time "may be due to changes in language and culture—an extreme example of this problem would be taking a test written in Middle English" (2, p. 330).

Smith's evidence for this supposed temporal change in the difficulty level of Stanford-Binet items is given in Tables 3 and 4 of her article. Table 3 shows the percent-passing Binet digit span items in the 1916, 1937, and 1960 normative samples, and Table 4 shows the mean percent of children passing all the Binet items at each year from 6 through 12 in the normative populations of 1937 and 1960. In both Tables 3 and 4 there is a significant decrease

---

in percent passing from 1937 to 1960. In connection with the Table 3 figures, Smith comments "The digit-span items were not changed in age placement level and, for some reason, have consistently received fewer correct answers over the years" (2, p. 323-324). In connection with Table 4, Smith writes "Note that the mean percent passing for each year is higher for the 1937 test than the 1960 test, which in fact is a finding that test scores are lower for groups with the same characteristics as the original normative sample, at a later point in time" (2, p. 330).

This conclusion is based on a simple error in Smith's understanding of the percent passing figures in her Tables 3 and 4. The percent passing figures simply are not comparable for 1937 and 1960. These figures were obtained from Appendix B in the Terman-Merrill Manual (3, pp. 342-347). In the Terman-Merrill table, the percent passing (each item) in 1937 is based on the percentage of children within each *chronological age (CA)* interval who pass the item; the percent passing in 1960, however, is based on the percentage of children within each *mental age (MA)* interval who pass the item. This important distinction is explained by Terman and Merrill (3, pp. 28-31), who take pains to point out that the discrepancy between the 1937 and 1966 percents passing are an inevitable and predictable result of the fact that the one is based on *CA* groups and the other on *MA* groups. They give a specific example why this is so. Referring to the Vocabulary test at the VIII-year level, "we note that no child who has a mental age of 5 passes the test whereas 4 of the brightest 5-year-olds pass this VIII-year level test. Seventeen of the bright 6-year-olds pass it, but only 5 whose mental age is 6 succeed. The mental age group fails to include those brighter 5- and 6-year-olds who pass this test, but who, because they have higher mental age scores, do not fall into the V and VI-year mental age groups" (p. 30-31). They point out that the differences in the 1937 and 1960 percents passing (based on *CA*s and *MA*s, respectively) "does not mean that in 1960 the Vocabulary test at this level is failing to maintain the 1937 level of difficulty."

Thus, *the temporal differences in difficulty level of Binet items reported by Smith are merely spurious with respect to her hypothesis. In no way can they support her conclusion of a temporal cultural bias in the Stanford-Binet test.*

The same error is repeated in Smith's Figures 3, 4, and 5 which graphically plot the 1937 and 1960 Terman normative data.

Another similar error is seen in Smith's listing (in her Tables 3 and 4) of percent passing at each age by U.S. blacks. The reader would assume that the percent passing at each age was determined in the same way for the 1937

and 1960 Terman U.S. norms and for the U.S. black norms (1). But the percent passing on the black norms are based on children grouped by *grade level* in school. Over-age duller children will more likely be in lower grades for their *CA;* under-age brighter children will more likely be in higher grades for their *CA*. So the percentage passing (a test item) based on grouping by school grade cannot properly be compared to a percentage passing based on grouping by *CA*, which, finally, cannot be compared with grouping by *MA*. Yet Smith presents the three columns of figures alongside one another and makes comparisons and interpretations as if all three sets of figures were comparable measures! Earlier in her article, Smith herself notes that "the data for the Kennedy standardization for southern United States blacks are not adequate for this study. The data for the normative sample are given for school grades one through six, not by age" (2, p. 318).

Finally, serious doubts seem justified concerning an analysis of cultural biases in Binet items that reports quantitative results based on a handpicked selection of only eight items out of the total of 192 which make up the Stanford-Binet tests.

REFERENCES

1. KENNEDY, W. A., VAN DE RIET, V. & WHITE, J. C., JR. A normative sample of intelligence and achievement of Negro elementary school children in the Southeastern United States. *Monog. Soc. Res. Child Devel.*, 1963, **28** (No. 6).

2. SMITH, M. W. Alfred Binet's remarkable questions: A cross-national and cross-temporal analysis of the cultural biases built into the Stanford-Binet Intelligence Scale and other Binet tests. *Genet. Psychol. Monog.*, 1974, **89,** 307-334.

3. TERMAN, L., & MERRILL, M. A. Stanford-Binet, 1960, Intelligence Scale. Manual for the 3d Revision Form L-M. Boston: Houghton-Mifflin, 1960.

Institute of Human Learning
University of California
Berkeley, California 94720