Journal of Consulting Psychology
1967, Vol. 31, No. 6, 626–630

DRAW-A-PERSON TEST AS A MEASURE OF INTELLIGENCE IN PRESCHOOL CHILDREN FROM VERY LOW INCOME FAMILIES

LOIS-ELLIN DATTA[1]

National Institute of Mental Health

Ethnic group and sex did not affect either the IQ-equivalent scores or the congruent validities of the figure-drawing test (Goodenough-Harris scoring) in a nationally representative sample of 956 children enrolled in full-year Head Start programs. Higher correlations were obtained for performance (Caldwell-Soule Preschool Inventory) than for verbal (PPVT) measures and for older (4–7 yr.) than for younger children. The obtained coefficients (.3–.5) compare favorably with those previously reported for kindergarten and 1st-grade children from less impoverished families. Both PPVT and DAP mean IQ equivalents indicated, however, substantially lower performance for Head Start than for normative groups.

Recent interest in special educational programs has drawn attention to problems in measuring intellectual abilities and changes in performance in preschool children from low income families. The question of "culture-fairness" is thus added to the already considerable task of obtaining reliable measures at an age when the behavioral repertoire is limited.

The figure-drawing test has been widely used as a measure of intelligence in children (Sundberg, 1960). It is simple to administer and score and is considered to have predictive and congruent validity coefficients that, while relatively low, compare favorably with those reported for other standardized intelligence and achievement tests (Shipp & Loudon, 1964; Vane & Kessler, 1964). Dennis (1966) concluded that Draw-A-Person performance reflects experience with representational art rather than parental education or literacy.

The availability of data from a nationwide sample of children enrolled in Project Head Start centers provided an opportunity to estimate the congruent validity of the figure-drawing test for younger children from very low income families. This was measured by comparing the results of the Draw-A-Person test (DAP) with the results obtained from the Peabody Picture Vocabulary Test (PPVT) and the Caldwell-Soule Preschool Inventory (PSI).[2] The PPVT (Dunn, 1965) is a widely used measure of verbal intelligence; the PSI has been developed as a "culture-fair" measure of intelligence in preschool children. The DAP requires less equipment, administration time, and examiner training than does the PPVT. The PSI is similar to the WISC in terms of equipment, examiner training, administration and scoring time, and the apparent contribution of verbal and nonverbal skills to test performance.

For the DAP, psychometrically desirable characteristics of a culture-fair test would include (*a*) a mean standard score of about 100 and (*b*) correlations between the DAP and the PPVT and the DAP and the PSI at least similar in magnitude to validity coefficients typically reported for the DAP (Harris, 1963).

Method

Seventy-two Project Head Start centers were selected to provide a sample representative of the population of 1966 full-year program centers in terms of geographic distribution and program length. From each center, 12–15 children were selected at random from an identification number list for inclusion in the survey.

The DAP, PPVT, and PSI were administered individually by college graduates with special training

[1] This study used data collected by the Planning Research Corporation for the Office of Economic Opportunity under Contract No. OEO-1308, 1966. The author wishes to thank Ann Drake, for her assistance in data analysis, and Ruth Ann O'Keefe, for her contributions to every phase of the study.

[2] B. Caldwell and D. Soule, The Preschool Inventory. Unpublished paper, Project Head Start, Office of Economic Opportunity, Contract S14, 1966.

in examining disadvantaged children. Sixty-five children, who were predominantly of Mexican origin, were tested in Spanish. Since the equivalence of the Spanish and English versions of the PPVT was not determined, data from children tested in Spanish were analyzed separately from data of children tested in English.

Draw-A-Person (Machover, 1948) rather than Draw-A-Man instructions were used for the figure-drawing test. Data on sexual identification will be reported in a later paper. Bliss and Berger (1954) have concluded that the two forms of the test yield substantially the same results. Unless the drawing was identified as a woman by the child, ambiguous figures were scored by the Goodenough-Harris criteria for drawings of men (Harris, 1963). Of the 956 drawings, 239 were not recognizable figures (Class A), 111 were scored by Draw-A-Woman criteria, and 606 were scored by Draw-A-Man criteria. Interrater reliabilities among the four scorers ranged .89–.99 for samples of 14–50 drawings. (For a detailed report of sampling, selection, and testing procedures, see Commins, Cort, Henderson, & O'Keefe, 1967.)

RESULTS AND DISCUSSION

Mean Standard Scores

The DAP and PPVT raw scores were converted to standard scores; the mean standard score at each age is set at 100 for the normative samples for both tests. Table 1 shows that regardless of age, sex, or ethnic group, the average performance on both the DAP (overall mean standard score, 77.22) and the PPVT (overall mean standard score, 82.02) was substantially lower than the mean for the normative samples.

The low PPVT standard scores are consistent with the poor performance on verbal tasks frequently reported for children from lower class and minority group families (Deutsch, 1965). The low DAP standard scores were to some extent unexpected. Previous studies have indicated that at least four relatively disadvantaged groups have achieved mean standard scores of about 100 on the DAP. Such means were reported for white and Negro kindergarten children in New York City public schools (Vane & Kessler, 1964), for white and Negro 5-year-old children in a New York City day care center (Anastasi & D'Angelo, 1952), and for a representative sample of 300 Negro first-grade children from southeastern states (Kennedy & Lindner, 1964). Bowers and Giles (1966) found an increase in DAP scores as socioeconomic status increased among 6- to 12-year-old children in Evanston, Illinois, but the mean DAP standard scores for the lowest socioeconomic groups, regardless of sex or age, were about 100.

The norms for younger children on the DAP are not geographically and economically representative of the national population at

TABLE 1

DAP, PPVT, AND PSI CORRELATIONS BY SEX, ETHNIC GROUP, AND AGE IN A SAMPLE OF 956 PRESCHOOL CHILDREN ENROLLED IN PROJECT HEAD START

Group	*N*	*M* age[a]	*M* standard score		Raw-score correlation		
			PPVT	DAP	DAP/PPVT	DAP/PSI	PPVT/PSI
Total	956	60.3	82.02	77.22	.46	.56	.73
Sex, ethnic group[b]							
Boys, white	188	63.3	89.26	76.57	.40	.51	.69
Boys, Negro	273	56.4	80.64	76.02	.52	.56	.72
Girls, white	166	62.1	85.18	78.14	.51	.60	.77
Girls, Negro	264	58.2	79.32	77.45	.44	.54	.65
Spanish-speaking	65	71.8	69.77	80.66	.53	.60	.72
Age[b]							
3	72	44.1	80.01	76.76	.38	.39	.48
4	397	54.9	82.23	78.76	.22	.26	.69
5	335	63.8	84.32	74.21	.31	.44	.62
6	87	76.0	83.02	79.53	.52	.57	.80

[a] In months.
[b] Includes only children tested in English; too few children were tested in Spanish to compute data for sex and age subgroups.

these ages; the present sample differs from the normative samples and from the four cited lower income groups in geographic distribution and in degree of economic deprivation. Eisenberg and Conners (1966) have reported a DAP mean standard score of about 81 for 712 children entering Baltimore Head Start classes; of these children, 44% came from families with an annual income of less than $3,000, 30% were supported by welfare, 64% of the fathers were unskilled, and about 60% of both parents had less than a tenth-grade education. In the present sample, the mean annual income was $3,771 for a living group which averaged 6.7 persons, approximately $560 per year per person.

In comparison, the median educational level completed by the parents of the New York City children (Anastasi & D'Angelo, 1952) was the eleventh grade, all of the mothers were employed, the median number of siblings was 1, and only 13% of the fathers were unskilled. The Baltimore children thus appear to differ from the New York City children primarily in the severity of economic deprivation; they appear to differ from the present sample primarily in geographic location and urban/rural distribution. Although the data do not permit estimations of the independent contributions of income, city size, geographic area, or of their interactions, factors associated with very low income rather than factors associated with geographic or urban/rural distributions seem to be responsible for the differences in DAP performance between Head Start children and the previously reported samples.

The low mean DAP standard scores indicate that, with the present norms, the test would not provide a "culture-fair" measure of individual attainment in an economically heterogeneous group. Within the Head Start sample, however, DAP standard scores were not affected by ethnic or sex differences, while the PPVT standard scores were affected by factors associated with both sex and ethnic group. Results of a 2×2 unweighted means analysis of variance (Winer, 1962) indicated that the DAP standard scores of girls and boys and of Negro and white children did not differ significantly. Ethnic group and sex F ratios significant at $\leqq .01$ were found for PPVT standard scores: The PPVT performance of white children was higher than the performance of Negro children ($F = 31.43$, $p \leqq .001$), and boys, regardless of ethnic group, achieved higher PPVT scores than did girls ($F = 4.37$, $p \leqq .01$). The PPVT mean standard score for Spanish-speaking children (69.77) was significantly lower than the PPVT means for other groups; the DAP mean standard score for Spanish-speaking children (80.66) did not differ significantly from the DAP mean scores for other groups. The DAP may thus be relatively insensitive to factors affecting the PPVT scores, and among these factors may be those related to cultural influences.

It is difficult to estimate the extent to which the low mean standard scores on both the DAP and the PPVT are due to cognitive as contrasted to emotional or motivational associates of deprivation. Some evidence of the importance of cognitive factors may be found in the report that culturally deprived children were not reliably lower on *all* measured aspects of psycholinguistic functioning, but were primarily handicapped in the areas of auditory word comprehension and auditory vocal automatic decoding (Barrett, Semmel, & Weener, 1965). On the other hand, relatively minor changes in testing conditions have been associated with substantial improvement in performance (Riessman, 1962). Despite agreement on the importance of optimum testing conditions, there have been few systematic studies comparing directive ("Think again; you can do better than that"), standard-neutral, and supportive attitudes for deprived and privileged preschool and elementary children.

Congruent Validity

The raw-score product-moment correlations among the DAP, PPVT, and PSI shown in Table 1 are all significant at $\leqq .01$; t comparisons among the correlations (r to z transformations) indicated that age, sex, and ethnic group did not significantly affect the congruent validity of the DAP. The correlations for all subgroups compare favorably with the .4 typically reported for groups of about 100 normal kindergarten and first-

grade children and with the .39 DAP/PPVT raw-score correlation obtained in a sample of 5- to 6-year-old Head Start children (Eisenberg & Conners, 1966). The DAP/PPVT standard-score product-moment correlations ranged .19–.52; these correlations, while lower than the raw-score correlations, were significant at $\leqq .01$. The single exception was .22 obtained for 3-year-old children.

Test bias has been defined (Educational Testing Service, 1966) as the consistent over- or underprediction of a criterion in one subgroup as compared to another subgroup, so that equally high predictive validity within subgroups would indicate a lack of bias. If this definition is extended to congruent validity, the value of the DAP as an estimate of general intelligence appears to be as high among children between 4 years, 0 months and 6 years, 11 months from very low income families as it is among children of this age or slightly older from less economically deprived backgrounds.

Age and validity. The validity of the DAP for school children has previously been reported to decrease with age, being higher for children in kindergarten and the first grade than for children older than 9 years (Ellis, 1953; Kennedy & Lindner, 1964; Pringle & Pickup, 1963; Vane & Kessler, 1964). As sample size increases, correlation magnitudes tend to decrease. If sample size is considered in this preschool sample, the DAP/PPVT raw-score correlations tend to increase with age. The lower congruent validity of the DAP for the younger children suggests that the value of the DAP as a measure of intelligence in children may be curvilinear with respect to age, increasing from 3–5 years and decreasing after about 8 years of age.

Performance and verbal measures. The congruent validity of the DAP was higher for performance (.56) than for verbal (.45) abilities, regardless of age, sex, or ethnic group. Similar results have been reported by Pringle and Pickup (1963) and Harris (1959). These correlations were, however, considerably lower than the PPVT/PSI correlation of .73; the PSI would appear to have more reliable variance associated with a verbal than with a performance measure of intelligence.

SUMMARY AND CONCLUSIONS

The DAP, the PPVT, and the PSI were administered to a nationally representative sample of 956 children attending 1966 Project Head Start full-year classes. Among children from 4 years, 0 months to 6 years, 11 months, the DAP/PPVT and DAP/PSI correlations compared favorably with validity coefficients previously reported for children from less deprived homes. Among younger children, the congruent validity of the DAP was lower. Neither sex nor ethnic group significantly affected DAP correlations; the DAP thus meets one criterion of a culture-fair measure.

On both the PPVT and the DAP, however, the mean standard scores were substantially lower than those reported for the norm groups. By the second criterion, the value of the DAP as a culture-fair measure of intelligence remains in question for children in samples heterogeneous for socioeconomic status, although within this very low income sample, the DAP was less affected than the PPVT by factors associated with ethnic group and sex.

REFERENCES

ANASTASI, A., & D'ANGELO, R. A comparison of Negro and white preschool children in language development and Goodenough Draw-A-Man IQ. *Journal of Genetic Psychology,* 1952, 81, 147–165.

BARRETT, L. S., SEMMEL, M. I., & WEENER, P. D. A comparison of the psycholinguistic functioning of "educationally deprived" and "educationally advantaged" children. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1965.

BLISS, M., & BERGER, A. Measurement of mental age as indicated by the male figure drawings of the mentally subnormal using Goodenough and Machover instructions. *American Journal of Mental Deficiency,* 1954, 59, 73–79.

BOWERS, N. D., & GILES, G. C. Social class, sex and age differences in sex role identity as inferred from children's drawings of men and women. Paper presented at the meeting of the American Psychological Association, New York, September 1966.

COMMINS, W. D., CORT, H. R., HENDERSON, N. H., & O'KEEFE, R. A. *A study of the 1966 full-year Head Start programs.* Preliminary Report D1268, 1967, Planning Research Corporation, Contract OEO-1308, Office of Economic Opportunity.

DENNIS, W. Goodenough scores, art experience, and moderation. *Journal of Social Psychology,* 1966, 68, 211–228.

DEUTSCH, M. The role of social class in language development and cognition. *American Journal of Orthopsychiatry,* 1965, **35**, 78–88.

DUNN, L. M. *Peabody Picture Vocabulary Test: An expanded manual.* Minneapolis: American Guidance Series, 1965.

EDUCATIONAL TESTING SERVICE. Are aptitude tests unfair to Negroes? ETS investigates two kinds of bias. *Developments,* 1966, **14**, 1–4.

EISENBERG, L., & CONNERS, C. K. The effect of Head Start on developmental processes. Paper presented at the 1966 Joseph P. Kennedy, Jr., Foundation Scientific Symposium on Mental Retardation, Boston, April 1966.

ELLIS, R. Comparison of scores on Goodenough Draw-A-Man, revised Stanford-Binet and Wechsler Intelligence Scale for Children, obtained from a group of children seen at a psychiatric clinic. Unpublished master's colloquium paper, University of Minnesota, 1953.

HARRIS, D. B. A note on some ability correlates of the Raven Progressive Matrices (1947) in the kindergarten. *Journal of Educational Psychology,* 1959, **50**, 228–229.

HARRIS, D. B. *Children's drawings as measures of intellectual maturity.* New York: Harcourt, Brace & World, 1963.

KENNEDY, W. A., & LINDNER, R. S. A normative study of the Goodenough Draw-A-Man test on southeastern Negro elementary school children. *Child Development,* 1964, **35**, 33–62.

MACHOVER, K. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1948.

PRINGLE, M. L., & PICKUP, K. T. The reliability and validity of the Goodenough Draw-A-Man test. *British Journal of Educational Psychology,* 1963, **33**, 297–306.

SHIPP, D. E., & LOUDON, M. L. The Draw-A-Man test and achievement in the first grade. *Journal of Educational Research,* 1964, **57**, 918–921.

SUNDBERG, N. Clinical assessment problems and procedures: A national survey. *American Psychologist,* 1960, **15**, 404.

RIESSMAN, F. *The culturally deprived child.* New York: Harper & Row, 1962.

VANE, J., & KESSLER, R. The Goodenough Draw-A-Man test: Long term reliability and validity. *Journal of Clinical Psychology,* 1964, **20**, 487–488.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received April 3, 1967)