R E P O R T R E S U M E S

ED 013 474 CG 000 931

THE RELIABILITY AND VALIDITY OF THE NEW HARRIS-GOODENOUGH DRAW-A-MAN TEST.

BY- DUNN, JAMES A.
MICHIGAN UNIV., ANN ARBOR,MIDWEST RESEARCH CTR.

PUB DATE 67

EDRS PRICE MF-$0.25 HC-$0.64 16P.

DESCRIPTORS- *TEST VALIDITY, *TEST RELIABILITY, ELEMENTARY SCHOOL STUDENTS, *SCORING, INTELLIGENCE TESTS, ACHIEVEMENT TESTS, STANFORD BINET INTELLIGENCE TEST, WECHSLER INTELLIGENCE SCALE FOR CHILDREN, IOWA TESTS OF BASIC SKILLS, CALIFORNIA TEST OF MENTAL MATURITY, HARRIS GOODENOUGH DRAW A MAN TEST

A SERIES OF STUDIES SOUGHT TO ASCERTAIN (1) INTER- AND INTRA-RATER RELIABILITY FOR THE HARRIS-GOODENOUGH DRAW-A-MAN TEST, (2) DRAW-A-MAN (DAM) VALIDITY, AND DRAW-A-WOMAN (DAW) VALIDITY. THE DRAWINGS OF 72 STUDENTS IN AN UPPER MIDDLE CLASS WERE SCORED INDEPENDENTLY BY TWO SELF-TAUGHT SCORERS, ONE OF WHOM ALSO RESCORED THE TEST A WEEK LATER. THE INTER-RATER RELIABILITY WAS .88 AND THE INTRA-RATER RELIABILITY WAS .93. IT SEEMS THAT NO SIGNIFICANT INCREASE IN RATER RELIABILITY WAS ACHIEVED BY HARRIS' REVISION. THE THREE DAM VALIDATION STUDIES RELATED DAM SCORES TO/ (1) STANFORD-BINET FORM L-M SCORES, (2) THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN, AND (3) GROUP MEASURES OF INTELLIGENCE AND ACHIEVEMENT. RESULTS ARE PRESENTED IN TWO TABLES. ALTHOUGH MODERATE CORRELATIONS WITH THE BINET AND WECHSLER SCALES WERE FOUND, THE TEST APPEARS TO TAP AREAS OF INTELLECTUAL ABILITY THAT HAVE LITTLE SIGNIFICANCE FOR ACADEMIC ACHIEVEMENT. IN THIS STUDY, LITTLE DIFFERENCE WAS FOUND BETWEEN THE DAM-WISC AND DAW-WISC CORRELATIONAL PATTERNS, SUGGESTING THAT THE TWO FORMS OF THE TEST ARE SIMILAR AND MIGHT BE USED INTER-CHANGEABLY IN DETERMINING MEAN GROUP IQ LEVELS. THIS DOCUMENT APPEARED AS STUDY 7 IN SCHOOL ANXIETY AND COGNITIVE FUNCTIONING/ EXPLORATORY STUDIES, REPORT 4, IRCOPPS MIDWEST RESEARCH CENTER FOR PUPIL PERSONNEL SERVICES, ANN ARBOR, MICH., PP. 212-222. (PS)

INTERPROFESSIONAL

RESEARCH

COMMISSION

ON

PUPIL

PERSONNEL

SERVICES

Midwest Research Center

SCHOOL ANXIETY AND COGNITIVE FUNCTIONING: EXPLORATORY STUDIES

University of Michigan

CG 000 931

Report 4

IRCOPPS MIDWEST RESEARCH CENTER
FOR PUPIL PERSONNEL SERVICES

SCHOOL ANXIETY AND COGNITIVE FUNCTIONING: EXPLORATORY STUDIES

James Dunn
Philip Safford
Ruth Schelkun
Roger Scott
Patricia Shanks

THE UNIVERSITY OF MICHIGAN
1967

MIDWEST RESEARCH CENTER FOR PUPIL PERSONNEL SERVICE

July 1967

PREFACE

This report is the fourth in a series of research monographs published by the IRCOPPS Midwest Research Center. A survey of Center activities plus a comprehensive synopsis of the Center's project reports may be found in the Center's 1967 Summary Status Report.

The present monograph reports the results of eight modular pilot studies conducted by various center staff. All research was supported by NIMH Grant #01428. Several of the studies have been presented, in abbreviated form, at various professional meetings and certain of the results have already appeared, or are due to appear, as short published articles.

Appreciation is expressed to the various staff associated with the production of these reports.

James A. Dunn
Director
IRCOPPS
Midwest Research Center

TABLE OF CONTENTS

STUDY VII

THE RELIABILITY AND VALIDITY OF THE NEW HARRIS-GOODENOUGH DRAW-A-MAN TEST

JAMES A. DUNN

TABLE OF CONTENTS

Human figure drawing is one of the most widely used assessment procedures in psychology. According to a study by Sundberg (1961) the Goodenough Draw-A-Man (DAM) Test was used in over 80% of the hospitals and institutions surveyed and it ranked just after the Wechsler-Bellevue Intelligence Test in over-all frequency of use. Until Harris' revision (1963), the Goodenough Draw-A-Man Test had not been changed since its publication (1926). Consequently Harris' revision has met with considerable interest in the field.

The major characteristics of the Harris revision are:

1) a more extensive, and presumably more objective, scoring system;

2) the utilization of the deviation rather than the mental age IQ concept; and

3) the development of an alternative, Draw-A-Woman (DAW), form of the test.

Although Harris based his revision of the scoring procedure on data obtained from a sample of 3000 carefully selected children, relatively little attention was paid directly to the question of reliability and validity. Most of the discussion of reliability and validity in Chapter 5 of Harris' book pertained to the 1926 scoring procedure.

Only two of the reliability studies reported by Harris were conducted since 1947. Both studies were conducted by Harris and although he did not specifically say the two studies were based on the revised scoring system, it would only be fair to assume that they were.

The more extensive of the two studies was based on 300 children of two age levels, 8 and 10. Inter-rater reliabilities for scoring the protocols of these children ranged from .92 to .98 depending on the age and sex of the subjects. No discussion was given of intra-rater reliability (i.e., consistency of same rater scoring) or test-retest reliability.

The second study was based on drawings of approximately 100 kindergarten children. Tests were given on each of ten consecutive days. Using analysis of variance, Harris reported no significant intra-child variation in Draw-A-Man scores. Such a prodecure does not provide information about the degree of relationship between various test scores of an individual, however; it only indicates that the differences between an individual's several scores are not so great but that they could be attributed to chance.

Regarding validity, none of the studies cited by Harris (1963) relating the Draw-A-Man Test to individual criterion tests, such as the Wechsler or the Binet, were published after 1953. Thus, there may also be some question regarding the validity of the new scale. Harris does indicate, however, that in a study based on the responses of 200 Canadian Indian

children, IQ's computed according to the 1926 procedure correlated .96 to .98 with IQ's computed according to the 1963 scoring procedure.

THE PROBLEM

Thus, in view of the relative sparcity of reliability and validity data on the new scoring system a series of studies was undertaken to ascertain: a) inter- and intra-rater reliability, b) DAM validity, and c) DAW validity.

Inter- and Intra-Rater Reliability

The Draw-A-Man test was given to all the children in Grades 1 through 6 in an upper middle class school. Twelve drawings, 6 for boys and 6 for girls, were selected at random from the data pool for each of the grade levels. Thus, the sample was stratified and balanced for grade and sex. Total number of S's was 72. The drawings were scored independently by two self-taught scorers. One week following initial scoring, one rater then rescored the protocols. Pearson product-moment correlations were then computed for the two sets of scores.

Results. The correlation between the scores produced by the two raters (inter-rater reliability) was .88 which is significantly different from zero ($p \leq .01$) and that between the first and second sets of scores produced by the same rater (intra-rater reliability) was .93. These values are almost identical with those given for the original (1926) Goodenough

scoring method by McCarthy (1944), who reported inter- and intra-rater reliabilities of .90 and .94, respectively. It seems that in spite of the greater length of the Harris procedure, and presumably its greater objectivity, there has been no significant increase in rater reliability achieved by the revision. The relative clarity of Harris' scoring procedure, however, may make it somewhat easier for an individual to achieve self-taught competence in scoring.

The DAM Validity Studies

Three separate DAM validation studies were conducted. The first related Draw-A-Man scores to Stanford-Binet Form L-M scores; the second related Draw-A-Man scores to scores on the Wechsler Intelligence Scale for Children (WISC); and the third related Draw-A-Man scores to group measures of intelligence and academic achievement, i.e., to California Test of Mental Maturity scores (CTMM) and to Iowa Test of Basic Skills (ITBS) scores.

In the first study the Binet and Draw-A-Man tests were administered to a sample of 32 presumably normal children randomly selected from a suburban elementary school; S's ranged in age from 6 to 10 years (M_{age} = 8.3; M_{IQ} = 107).

In the second study the WISC and the Draw-A-Man tests were administered to 93 randomly selected public school children, ranging in age from 6 to 15 (M = 10.7). The mean and standard deviation of the WISC IQ for this group were 100 and 23, respectively.

In the third study the Draw-A-Man, the California Test of Mental Maturity and the Iowa Test of Basic Skills were administered to 90 suburban elementary school children. Fifteen S's were randomly selected from each of the 6 elementary school grades.

All S's were drawn from public schools in middle class residential suburbs of a large Midwestern city. Each study was done independently of the others. Approximately equal numbers of males and females were used in each of the three studies. Pearson product-moment correlation matrixes were computed for each set of data.

Results. The correlation between Binet IQ's and Draw-A-Man IQ's was .78 ($p < .01$).

The correlations of Draw-A-Man scores and various Wechsler scores are presented in Table 1. These coefficients are, in general, somewhat higher than those ordinarily reported using the 1926 scoring procedure and are similar to those reported in an unpublished study done by Sister Mary Hilda in 1964.[2] Sister Hilda found that Draw-A-Man scores correlated .52 with IQ's on Form L-M of the Stanford-Binet and .37 with Quick Test scores. (The Quick Test is a non-verbal comprehension test of intelligence developed by Ammons and Ammons, 1962a, 1962b.)

---

[2]Sister Mary Hilda, S.C.C., A study of the inter-correlations of the Quick Test, Draw-A-Man Test, and the Stanford-Binet Intelligence Test, Form L-M. (Unpublished manuscript, Wayne State Univer., 1964)

TABLE 1

Product-Moment Correlations* Between Scores on Draw-A-Man and WISC Intelligence Scales (N = 32)

Full Scale IQ	.64	Vocabulary	.52
Verbal Scale IQ	.59	Picture Completion	.48
Performance IQ	.62	Picture Arrangement	.48
Information	.54	Block Design	.60
Comprehension	.53	Object Assembly	.52
Arithmetic	.49	Coding	.28
Similarities	.54		

*All rs significant at less than the .01 level.

Pearsonian correlations for Draw-A-Man (DAM) scores with CTMM scores and ITBS scores are summarized in Table 2. The DAM correlates .32 with CTMM ($p = .01$). With the exception of Reading Comprehension, correlations are uniformly very small and nonsignificant.

TABLE 2

Product-Moment Correlations Between Scores on Draw-A-Man, CTMM, and ITBS (N = 90)

CTMM Verbal IQ	.32*	ITBS Arithmetic	-.05
CTMM Non-verbal IQ	.17	ITBS Spelling	.06
ITBS Reading Comprehension	.20	ITBS Language Skills	.03

*$p < .01$.

On the basis of these results, it may be concluded that DAM IQ's derived from the new scoring procedure correlate somewhat better with individual tests of intelligence such as the WISC and the 1962 Stanford-Binet than did scores derived by the 1926 system and reported elsewhere. For the present groups, rs = .64 and .78, respectively. It is possible that this improvement in correlation may have been due, in part, to the use of deviation scores rather than mental-age-based IQ's.

There may be some question, however, about what the Goodenough test actually measures. While the test does have moderate correlations with such well accepted measures as the Binet and Wechsler scales, the test appears to tap areas of intellectual ability that have little significance for academic achievement. Correlations with verbal abilities are consistently lower than correlations with skills such as Block Designs and spatial perception. In view of this, it is suggested that the test might reflect such attributes and abilities as degree of concrete awareness, ability to comprehend social situations, and perhaps, even more fundamentally, the ability to develop and utilize concrete functional-motoric concepts as contrasted with abstract-verbal concepts.

The DAW Validity Study

Harris, in his revision, not only redeveloped and re-standardized the scoring procedure for the Draw-A-Man Test, he also included an alternate Draw-A-Woman (DAW) form. Inas-

much as the DAW is an entirely new instrument, there have, as yet, been no reports of correlations with criterion tests. Harris does claim reliability for the DAW and has reported DAW-Draw-A-Man correlations of .75.

In the present study, the Draw-A-Man and Draw-A-Woman tests and the Wechsler Intelligence Scale for Children were administered to 20, presumably normal, elementary school children in a middle class suburban school system. Their average age was 9.9. Mean WISC Full Scale IQ was 102 (SD = 22.6).

The Pearson product-moment correlations between IQ's computed from the Draw-A-Woman and Draw-A-Man tests and various scores derived from the WISC are presented in Table 3. In general, the Draw-A-Man-WISC correlations are considerably higher than those reported in the previous section. There was very little difference between the Draw-A-Man-WISC and Draw-A-Woman-WISC correlational patterns in the present study, suggesting that the two forms of the test are indeed quite similar and might be used inter-changeably in determining mean group IQ levels. However, Harris has indicated that there were significant sex differences in the execution of the two drawing tests which would preclude their interchangeable use with individual S's.

TABLE 3

Draw-A-Woman, Draw-A-Man, and WISC Intercorrelations

Score	DAW	DAM
DAM IQ	.87	
WISC Full Scale IQ	.81	.77
WISC Verbal IQ	.77	.73
WISC Performance IQ	.79	.75
Information Scaled Score	.53*	.51*
Comprehension Scaled Score	.74	.74
Arithmetic Scaled Score	.64	.62
Similarities Scaled Score	.67	.53*
Vocabulary Scaled Score	.81	.73
Picture Completion Scaled Score	.59	.68
Picture Arrangement Scaled Score	.58	.51*
Block Design Scaled Score	.74	.74
Object Assembly Scaled Score	.79	.73
Coding Scaled Score	.58	.49*

*$p < .05$; other rs, $p \leq .01$.

SUMMARY

In summary then, inter- and intra-rater reliabilities of the DAM test of intelligence were .88 and .93, respectively. Correlations of DAM IQ scores with individually administered IQ scores are moderate to good (.64 to .78); but correlations with group administered IQ scores were somewhat poorer (.32). DAW scores correlate .87 with DAM scores, and .81 with WISC Full Scale scores.

REFERENCES

Ammons, R. B., and Ammons, C. H. Quick Test. Missoula, Mont.: Psychological Test Specialists, 1962. (a)

Ammons, R. B., and Ammons, C. H. The Quick Test: provisional manual. Psychol. Rep., 1962, 11, 111-161. (b)

Goodenough, F. L. Measurement of intelligence by drawings. New York: World Book, 1926.

Harris, D. B. Children's drawings as measures of intellectual maturity. Harcourt, Brace and World, Inc., 1963.

McCarthy, Dorothea. A study of the reliability of the Goodenough drawing test of intelligence. J. Psychol., 1944, 18, 201-206.

Sundberg, N. D. The practice of psychological testing in clinical services in the United States. Amer. Psychol., 16, 79-83.