Proceedings of the Iowa Academy of Science

Volume 52 | Annual Issue Article 30

1945

A Method for Evaluation and Equation of Test Forms

Mary Dodds
Iowa State College

Barbara Day
Iowa State College

A. R. Lauer
Iowa State College

A METHOD FOR EVALUATION AND EQUATION OF TEST FORMS

MARY DODDS, BARBARA DAY AND A. R. LAUER

PROBLEM

Notwithstanding the pioneer work of Spearman (1927(and numerous studies by Thurstone on problems of intelligence, most testing programs seem to postulate only one type of intelligence. Division of the total score may be made into arbitrary categories for use in advising students but the program itself does not seem to conform with accepted theory on the subject. There has been little change in the administration of the program.

This is partly due to the elaborate nature of test materials used. To save labor in developing a test it has been customary to build a long one to cover a given area most comprehensively. In many fields it has been found that less material carefully selected will do the job as well or better, thus affording notable economy in time and effort.

Further, there is a limit to the measurement of human capacities by the longer methods used in the past. In order to sample human traits adequately it is necessary to reduce the length of conventional tests without doing violence to the reliability of scores. This problem was studied by Rostron and Lauer (1939) and the ground work laid for a testing program of diversified nature possible of administration within limits of time normally available in crowded orientation programs.

The present study is an extension of the above paper designed to facilitate the development of parallel forms of tests. It is presented as a method of approach when limited time and energy are available to spend on the construction and application of measuring instruments.

METHOD AND PROCEDURE

Form A of the Iowa State Alertness Test was available in adequately standarized form. Its reliability and validity had been satisfactorily established. This test consisted of three sections, the first on arithmetic, the second on vocabulary and a third on types of relations. The problem set for experimental investigation was the development of parallel forms of tests having equal difficulty. This can be done a number of ways but too frequently the results finally obtained are not equivalent. The exact procedure followed will next be explained.

Two other forms of the test were made up consisting of equivalent items, as far as they could be judged from a priori evidence. The three forms A, B and C were then given to similar groups of 317, 132 and 122 cases respectively. An item analysis of the three sets of data was made and calculations of the number of items right, the

number wrong and the number omitted were made. Percentage indices were calculated for the number right and the number attempted—the latter proved highly variable and could not be used for evaluation or for scaling purposes.

Next, the sub-sections for each form of the test were compared by statistical techniques and the differences were evaluated. Where differences were greater than the one per cnt criterion, rearrangement and interchange of the items was made between the subjects of the different forms, using the percentage right as a criterion. This process was pursued until a balance of difficulty was obtained. See table II. These forms will ultimately be placed on electrical-scoring forms and re-standardized completely as separate units. Preliminary study of the results seems to warrant use of this method for the development of parallel forms of a given test.

RESULTS

After a complete analysis of all forms, the means and standard deviations were calculated and evaluations made by standard techniques described by Garret (1926). Table I gives the summary data on differences found between forms and sub-sections of forms.

Table I

Summary of Differences found between Sub-sections of Test Forms

Forms Compared	Test Elements	Means		Critical Ratio	Form most difficult
		A	B		
A and B	Arithmetical problems	3.60	3.37	2.08	Form B
		A	C		
A and C	Arithmetical problems	3.60	3.79	1.24	Form A
		B	C		
B and C	Arithmetical problems	3.37	3.79	2.95	Form B
		A	B		
A and B	Vocabulary	13.47	16.14	12.13	Form A
		A	C		
A and C	Vocabulary	13.47	17.25	18.35	Form A
		B	C		
B and C	Vocabulary	16.14	17.25	4.42	Form B
		A	B		
A and B	Relationships	4.25	5.32	5.85	Form A
		A	C		
A and C	Relationships	4.25	5.13	6.40	Form A
		B	C		
B and C	Relationships	5.32	5.13	1.10	Form C

It is shown here that in most respects Form A is more difficult than either Form B or C. In order to equate the forms treatment of each sub-section was made as shown in table II. In this way the forms are balanced as to difficulty and scaling is accomplished at the same time.

Table II

Final Regrouping of Items according to Per Cent Evaluation
Test 1, Forms A, B, and C

	Form A			Form B			Form C	
	Original Position	Per-centage		Original Position	Per-centage		Original Position	Per-centage
1.	A-2	82.9	1.	B-1	92.6	1.	B-2	82.2
2.	C-2	76.6	2.	C-1	87.8	2.	B-2	82.2
3.	A-3	70.0	3.	C-3	80.5	3.	B-3	65.6
4.	C-5	42.3	4.	C-6	39.2	4.	A-4	49.9
5.	B-4	37.2	5.	B-6	35.9	5.	C-4	40.7
6.	A-6	35.7	6.	B-5	11.2	6.	C-7	17.1
7.	A-5	12.9	7.	C-8	9.6	7.	A-7	10.4
8.	A-9	5.1	8.	B-7	5.2	8.	C-10	8.9
9.	B-8	3.7	9.	B-10	3.7	9.	A-8	8.2
10.	C-9	0.0	10.	B-4	0.0	10.	A-10	3.8
	Mean	36.64		Mean	36.57		Mean	36.60

While there may be slight variations due to position in the test, these difficulties will probably balance in the final evaluation procedure. Table II is merely illustrative of the treatment accorded each sub-divsion of the test.

SUMMARY AND CONCLUSIONS

1. The oligarchic doctrine of intelligence increases the number of specific capacities which must be sampled in a testing program. Most older tests are too long to give proper scope to the measurements. Abbreviated tests are needed.

2. A method is described for evaluating and compositing test forms when a certain amount of standardized material is available.

3. The method is llustrated by results obtained in the development of three forms of an alertness test.

IOWA STATE COLLEGE

REFERENCES CITED

Garret, H. E. Statistics in Psychology and Education Longmans, Green & Co. 1926 pp. 287.

Rostron, Charles R. and Lauer, A. R. The reliability of an abbreviated test of alertness proceedings Iowa Academy Sci. 1939 pp. 319-320.

Separman, Carl. The Abilities of Man. Macmillan, 1927.