Research Memorandum 67-2

A MODEL SAMPLING EXPERIMENT TO EVALUATE TWO METHODS OF TEST SELECTION

March 1967

U. S. Army
Behavioral Science Research Laboratory

Army Project Number
2J024701A723

Computerized Manpower Systems c-15

Research Memorandum 67-2

A MODEL SAMPLING EXPERIMENT TO EVALUATE TWO METHODS OF TEST SELECTION

by Robert N. Harris

Richard C. Sorenson, Task Leader

BESRL RM-67-2

Submitted by:
Cecil D. Johnson
Chief, Statistical Research
and Analysis Division

Approved by:
J. E. Uhlaner, Director
Behavioral Science
Research Laboratory

March 1967

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the U. S. Army Behavioral Science Research Laboratory.

402 177

A MODEL SAMPLING EXPERIMENT TO EVALUATE TWO METHODS OF TEST SELECTION

BACKGROUND AND PURPOSE

Assignment of Army enlisted personnel to Advanced Individual Training following basic training is based on a battery of tests called the Army Classification Battery (ACB). Currently, the ACB scores are combined in pairs to form aptitude area scores each of which is considered predictive of performance for the several Military Occupational Specialties (MOS) within one or more occupational categories. Optimal assignment of enlisted men to these specialties is based on these estimates of performance.

As part of the Army research program to improve assignment of personnel, new test batteries to complement or replace existing test batteries are continually being developed. A procedure generally used in the development of test batteries involves 1) the formation of an experimental battery composed of a large number of tests--the test pool--and 2) the selection of some subset of the pool to form an operational battery. Selection of tests to predict a single criterion of job performance is relatively simple and straightforward; tests are usually sequentially selected to maximize the multiple correlation coefficient under the constraint that all previously selected tests are retained. However, the selection of tests to predict several criteria can be accomplished by several different methods. The present Research Memorandum describes the use of model sampling techniques to evaluate two of these test selection methods.

TEST SELECTION PROCEDURES

The two test selection methods compared in this study were developed by Horst: 1) selection for absolute prediction, and 2) selection for differential prediction. The absolute method selects tests to maximize the average variance of the predicted criteria and is equivalent to maximizing the sum of the squared multiple correlation coefficients for predicting each criterion separately. The absolute method will yield the battery having the highest mean accuracy of prediction for all criteria. It is useful in many situations, as when personnel are selected if they exceed the cut-off score on one or more of the performance estimates and their assignment is determined from other considerations. The second method, for differential prediction, selects tests to maximize the average variance of the differences between all possible pairs of predicted criteria. When criteria and predictors are in standard form, this procedure is equivalent to maximizing the difference between the average variance of the predicted criteria and the average of their covariances. Batteries selected by this method will provide the best prediction of relative success in several jobs.

STANDARD OF COMPARISON

Which method is best depends on the objective of the test selection. The classical and most common standard of comparison is the validity of the battery when a single criterion is used (or the sum of the squared validity coefficients when multiple criteria are used). Since the absolute method selects tests specifically to maximize validity, the absolute method is, by definition, best by this criteria. No other method could possibly be superior. Similarly, the differential selection method would be best if evaluated against the function it maximizes. The function maximized by either method would obviously not provide a meaningful standard for comparison of the two selection methods.

The standard used in the present study was the extent to which the batteries accomplished the purpose for which they were constructed--to provide maximum improvement in the optimal assignment of personnel to jobs. The batteries selected by the two methods were therefore evaluated on the basis of average expected performance resulting from optimal assignment to jobs, with quota restrictions on the number assigned to each job.

VARIABLES

The group of tests comprising the test pool is listed in Table 1. The order of selection for the first 20 tests by each method is given in parenthesis. The criteria used for the test selection were school final course grades for the 12 MOS listed in Table 2.

PROCEDURES

This study was carried out in four steps: 1) selection of tests, 2) simulation of scores of the selected tests, 3) optimal assignment based on the simulated test scores, and 4) evaluation of the tests using results of optimal assignment. Each of these steps is discussed in turn.

A battery of 32 tests was used as the pool from which smaller batteries of tests were selected. Intercorrelations calculated from a sample of 2480 subjects were available, along with coefficients of validities for 12 groups of jobs, each based on different subjects. Sample size in the jobs ranged from 103 to 305. Absolute and differential test selection procedures were applied to the pool, and batteries of size 5, 10, and 20 tests were selected by each of the two methods. These six batteries are henceforth referred to as the selected batteries.

Table 1

BASIC SET OF TESTS AVAILABLE FOR SELECTION BY TWO METHODS[a]

Army Classification Battery

Verbal (20, 4)[b]	Shop Mechanics
Arithmetic Reasoning (1, 11)	Automotive Information (2, 7)
Pattern Analysis	Electronics Information (3, -)
Mechanical Aptitude	Classification Inventory (-, 2)
Army Clerical Speed (18, 17)	General Information (12, 9)
Army Radio Code (11, 20)	

Army Differential Aptitude Series

Object Completion (15, 14)	Letter Combinations
Word Squares	Hidden Figures
Pattern Analysis (-, 8)	Attention to Detail (10, 16)
Practical Situations	Patterns
Reaction to Signals	Perceptual Speed (6, 3)
Mechanical Principles (9, 18)	Associative Memory (14, -)
Spatial Orientation	Subtraction and Division (16, 10)

Noncognitive Scales

Clerk a priori (17, 15)	General Adjustment Empirical (19, 19)
Electronics a priori (8, 1)	Clerk Empirical (7, 12)
Mechanic a priori (5, 5)	Mechanic Empirical (13, 13)
Mechanic Suppressor (4, 6)	

[a] The test selection, done by Dr. William H. Helme, was reported in Technical Research Note 155, Evaluation of Differential Classification Tests for the ACB, June, 1965. U. S. Army Personnel Research Office, Washington, D. C. 20315.

[b] The first number in parenthesis is order of selection for absolute prediction; the 2nd number is order for differential prediction.

Table 2

CRITERION VARIABLES USED FOR TEST SELECTION

MOS No.	Approximate equivalent in current job structure	Title
223.1		Missile electronics mechanic
250.0		Electronics repair helper
271.1	32B	Fixed station receiver repairman
281.1	26L	Microwave radio repairman
293.1	31M	Radio relay and carrier operator
294.1	31L	Field carrier equipment repairman
296.1	31E	Field radio repairman
310.0	36A	Field communication crewman
321.1	36F	Lineman
352.1		Engineer missile equipment specialist
440.0	44A	Metalworking helper
551.1	51B	Carpenter
530.0	54A	Chemical operations helper
550.0	56A	Supply handler
626.1	62E	Construction machine operator
627.1	62F	Crane shovel operator
670.0	67A	Aircraft maintenance crewman
680.0	68A	Aircraft components repair helper
723.1	72B	Communications center specialist
724.1	72C	Switchboard operator
053.1	05C	Radio teletype operator

Samples of personnel were simulated using the assumption that the test scores could be represented by normal distributions. Hence, by model-sampling techniques, scores based on each of the six batteries were generated by computer to have the statistical characteristics of samples from a normal population. In addition, the covariance matrices of the normal populations used to generate the scores were equal to the covariance matrices of the six selected batteries. Either a 5- 10- or 20-variable normal distribution was used for the sample, depending on the size of the simulated selected battery. Ten samples of 216 subjects, each sample having different normally distributed scores were generated for each selected battery. Different normal variates (scores) were used for batteries of different size, but within each battery size the same normal variates were used for batteries selected by each method. Thus, the same simulated subjects were used in absolute and differential batteries of the same size.

Assignment of simulated subjects to jobs was based on estimates of performance for each subject in each job. The performance estimates were calculated separately for each selected battery from the simulated test scores by the least-squares regression equations. Uniform quotas were used in the assignment.

Assignment was evaluated using least squares regression estimates of performance, based on all 32 tests, in the job to which each subject was assigned. These performance estimates were averaged over subjects, yielding the assignment average. Since 10 samples were generated and assigned for each selected battery, 10 assignment averages were calculated for each battery. Each of the six sets of 10 assignment averages was then averaged to obtain six mean assignment averages; these mean assignment averages were used to compare the assignment effectiveness of the six selected batteries.

The differences between the mean assignment averages for the six selected batteries were tested for significance by the t-test for correlated means. The mean assignment averages and the t-test values are shown in Table 3. Under random allocation, the mean assignment averages would have a mean of 100 and a standard deviation of 20.

Table 3

MEAN ALLOCATION AVERAGES[a]

Battery Size	Absolute	Differential	t-value
5	109.79	110.89	6.45*
10	112.99	113.24	4.76*
20	113.99	114.15	9.76*

*Significant at the .01 level.

[a] The allocation average is the objective function for the linear programming model divided by the size of the sample on which objective function is maximized.

RESULTS

Examination of Table 3 shows that differential selection resulted in the more efficient assignment for all three battery sizes. This finding is interesting in view of the pattern of selected test: absolute and differential batteries of 5 tests had only 1 test in common (20%), while the batteries of 20 tests had 18 tests in common (90%). The difference between batteries of 20 tests was more significant than was the difference between the 5-test batteries in spite of the much greater overlap between 20-test batteries. Note the relationship between assignment efficiency and validity (Table 4 and 5). The absolute battery of 5 tests had substantially higher validity (for each MOS) than did the differential battery; validity coefficients for the 20-test batteries were very similar. Yet for both batteries the differential battery resulted in more efficient assignment.

In summary, when a test battery is to be used for optimal assignment of personnel, the evidence presented shows clearly the superiority of differential selection over absolute selection, at least for the variables used. Studies are now being conducted to establish the generality of this finding using test pools of several different sizes and several different numbers of criteria. Future studies will also use criteria covering another range of content, in an effort to establish the generality of the finding.

Table 4

VALIDITY COEFFICIENTS FOR BATTERIES OF 5 TESTS, 10 TESTS, 15 TESTS, AND 20 TESTS SELECTED BY THE ABSOLUTE METHOD[a]

Order of Selection	Selected Test	MOS: 223[c]	250	271	293	296	321	357	440	511	626	670	723
		Multiple Validity Coefficient[b]											
1	Arithmetic Reasoning (AR)	61	67	74	66	66	41	69	65	56	57	67	51
2	Automotive Information (AI)	63	68	80	69	73	58	72	70	63	66	75	51
3	Electronics Information (ELI)	64	71	83	71	75	59	76	71	63	67	76	52
4	Mechanic Suppressor (S-7)	74	71	84	71	76	59	78	71	63	68	76	52
5	Mechanic a priori (M-2)	74	73	84	71	76	63	82	72	65	68	76	52
6	Perceptual Speed (PS)												
7	Clerk empirical (C-7)												
8	Electronics a priori (E-2)												
9	Mechanical Principles (MP)												
10	Attention to Detail (AD)	80	76	90	75	79	68	90	80	67	72	78	54
11	Army Radio Code (ARC)												
12	General Information Test (GIT)												
13	Mechanic empirical (M-7)												
14	Associative Memory (AM)												
15[d]	Object Completion (OC)	85	76	94	81	80	71	91	83	71	74	79	58
16	Subtraction and Division (SD)												
17	Clerk a priori (C-2)												
18	Army Clerical Speed (ACS)												
19	General Adjustment empirical (G-7)												
20	Verbal (VE)	89	79	95	83	81	76	84	84	72	76	81	59
32	All Tests	90	79	95	84	82	76	94	85	72	77	81	60

[a]Op. cit., see Table 2.

[b]Decimal points omitted.

[c]MOS titles given in Table 2.

[d]A battery of 15 tests was not simulated; Validity coefficients for a 15-test battery are included as incidental information.

Table 5

VALIDITY COEFFICIENTS FOR BATTERIES OF 5 TESTS, 10 TESTS, 15 TESTS, AND 20 TESTS
SELECTED BY THE DIFFERENTIAL METHOD

Order of Selection	Selected Test	MOS: 223[c]	250	271	293	296	321	357	440	511	626	670	723
		Multiple Validity Coefficients[b]											
1	Electronics a priori (E-2)	32	32	50	26	40	20	43	33	28	13	29	10
2	Classification Inventory (CI)	33	40	53	48	50	33	49	54	44	37	41	33
3	Perceptual Speed (PS)	52	52	66	61	55	45	50	55	48	42	49	41
4	Verbal (VE)	64	65	78	69	65	46	64	69	55	53	68	50
5	Mechanic a priori (M-2)	66	66	78	69	66	59	66	71	55	59	70	50
6	Mechanic Supressor (S-7)												
7	*Automotive Information* (AI)												
8	Pattern Analysis (PA-X)												
9	General Information Test (GIT)												
10	Subtraction and Division (SD)	70	73	86	76	75	68	79	77	69	68	78	56
11	Arithmetic Reasoning (AR)												
12	Clerk empirical (C-7)												
13	Mechanic empirical (M-7)												
14	*Object Completion* (DC)												
15[d]	Clerk a priori (C-2)	86	76	92	79	80	70	88	82	[illegible]	70	79	57
16	Attention to Detail (AD)												
17	Army Clerical Speed (ACS)												
18	Mechanical Principles (MP)												
19	General Adjustment empirical (G-7)												
20	Army Radio Code (ARC)	89	79	94	81	81	75	91	85	72	76	81	60
32	All Tests	90	79	96	84	82	76	94	85	72	77	81	60

[a]Op. cit., see Table 2.

[b]Decimal points omitted.

[c]MOS titles given in Table 2.

[d]A battery of 15 tests was not simulated; Validity coefficients for a 15-test battery are included as incidental information.