

UNCLASSIFIED

AD NUMBER

ADB951523

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited. Document partially illegible.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors;
Administrative/Operational Use; 09 JUL 1942.
Other requests shall be referred to Office of the Adjutant General's Office (Army), Washington, DC 20310. Document partially illegible.

AUTHORITY

ARI ltr 13 Nov 1979

THIS PAGE IS UNCLASSIFIED

THIS REPORT HAS BEEN DELIMITED
AND CLEARED FOR PUBLIC RELEASE
UNDEF DOD DIRECTIVE 5200.20 AND
NO RESTRICTIONS ARE IMPOSED UPON
ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

UNANNOUNCED

119 Jul 42

12

17

AGO-354

T.D.A. Unit
JMC/dlk
7/9/42

AGO Rept. 350

6 Analysis of VC-1, X-1 Data from Camp Croft

MAR 6 1980

S
A
D

ADB951523

Problem

To analyze VC-1, X-1 scores of two groups of trainees, both groups being largely Grade V on GCT and 2abc, one group including men in regular Rifle Companies; the other, men in Special Training Battalions or recommended for Special Training. Problems to be considered are the selection of appropriate scoring formula; estimation of reliability of total test and separate pages; intercorrelation of battery of tests including VC-1 X-1, GCT, 2abc, DST-1, DST-2 and MA-1; and the validity of VC-1 items.

Population

Group A:

75 white trainees, Grade V on both GCT and 2abc, regular rifle companies, Camp Croft, S.C. These men, although Grade V, are making acceptable soldiers under the regular training program

Group B:

29 white trainees, Grades IV and V on GCT and 2abc in Special Training Battalion, Camp Croft, S.C. The majority of these men are illiterates or non-English. They are in the Special Training Battalion because of one or more of the following factors: illiteracy, limited intelligence, personality maladjustment.

48 white trainees, Grades IV and V on GCT and 2abc, recommended for Special Training Battalion, Camp Croft, S.C. These men are defined as men who would be in the Special Training Battalion if there were room for them.

Variables

1. VC-1, X-1: - visual classification test consisting of 67 5-alternative, multiple-choice items, of visual perception, abstraction, and paired comparison types. It was designed originally for use at induction centers to weed out those illiterates, non-English, and men of low mentality who would not be of value in the Army. The time limit for this test is 30 minutes.

2. GCT: - general intelligence test in spiral omnibus form consisting of 150 4-alternative multiple-choice items of vocabulary, arithmetic, and box-counting types. This test was designed for use at reception centers to classify men who read and write English. The time limit is 40 minutes.

3. 2abc: - non-language classification test for measuring the abilities of men who do not read English readily and of those in Grade V on GCT. It consists of three sub-tests: 70 block-counting items, 237 Symbol association items, and 87 design comparison items. The time limit for the total test is 20 minutes.

DDC FILE COPY

79 11 20 JCU 003650

THIS DOCUMENT IS BEST QUALITY PRACTICABLE.
THE COPY FURNISHED TO DDC CONTAINED A
SIGNIFICANT NUMBER OF PAGES WHICH DO NOT

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DDC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

4. DST-1: - a non-verbal test which attempts to measure non-verbal reasoning ability in low grade men. It consists of 30 items, each requiring the subject to draw the last figure necessary to complete a series of figures. The time limit is 10 minutes.

5. DST-2: - a test consisting of 30 items testing ability to follow simple verbal directions. This test was designed for use with low grade men. The time limit is 4 minutes.

6. MA-1: - a test designed to estimate chances of success in learning mechanical tasks, consisting of 45 items of each of the following types: Mechanical movements, surface development, and shop mathematics. The time limit for the total test is 45 minutes.

Conclusions

1. The VC-1, X-1 may be scored by the formula number right. It is not necessary to use the formula $R = \frac{1}{2}W$ since the scores by the two formulas are highly correlated ($r = .99$) and the relationship is linear.

2. The reliability of the total VC-1, X-1 test by the Kuder-Richardson formula #21 is .864 for the combined groups A and B. It is slightly higher for Group B alone, but is still lower than what would be desired for a test of this type. The split-half reliability raised by the Spearman-Brown formula, is .939 for the combined groups. This technique yields the highest estimate of reliability while the K-R #21 yields the most conservative estimate.

3. The reliabilities of the separate pages of the VC-1, X-1 are still lower, ranging from .514 to .707 and are too low to permit the use of a separate page as a shorter test for preliminary screening purposes.

4. The DST-1 is less reliable than the VC-1, X-1, the reliability by the Kuder-Richardson formula #21 for the total sample being .708.

5. The DST-2 is highly reliable in spite of its 4 minute time limit. The reliability for the entire group is .947. With respect to reliability then, the DST-2 would be preferable to the VC-1, X-1, especially since it is so much shorter.

6. The intercorrelations of the battery of tests, including VC-1 X-1, GCT, 2abc, DST-1, DST-2 and MA-1, are very low with none exceeding .475. The VC-1, X-1 correlates most highly with the 2abc, DST-1, and DST-2, the correlations ranging from .43 to .46 for the total group. All four of these tests were specially designed for use with men who are Grade V on the verbal intelligence tests. The correlations of these tests with the GCT are practically zero for Group A and from .24 to .37 for Groups A and B combined. The intercorrelations involving DST-2 are consistently higher than the others.

7. The item analysis of VC-1, X-1 indicates that few of the items in this form of the test are sufficiently difficult or discriminating. The perception items are generally too easy; the analogy items, non-discriminating. A new form of the test should be constructed using the abstraction type of item, and a higher level of difficulty.

Procedure

The VC-1, X-1 test papers for the three groups of trainees were scored by two scoring formulas, number right, and right minus one-fourth wrong. Scores of the three groups were compared and the significance of the differences between the means obtained. It was decided to combine those recommended for Special Training with those in Special Training, calling this Group B; and calling those in Rifle Training, Group A.

In order to see if it would be worthwhile to use the $R - \frac{1}{4}W$ * scoring formula, scatter plots were made plotting R against $R - \frac{1}{4}W$ scores for Groups A & B separately; Correlations were obtained and the distributions checked for linearity. Since the correlations were very high and the linearity obvious it was decided that the $R - \frac{1}{4}W$ scoring formula could be discarded.

For R scores only, estimates of the reliability of the VC-1, X-1 were obtained by the following methods for Groups A & B separately and combined.

1. Kuder-Richardson formula #21, for total test.
2. Split-half method, odds vs. evens for total test.
3. Kuder-Richardson formula #21 for each page separately.

Estimates of the reliability of the DST-1 and DST-2 were also obtained by K-R #21.

Intercorrelations of the tests included in the battery were obtained, using all available cases for each correlation, for Groups A and B, separately and in combination.

An item analysis was made of VC-1, X-1, using the external criterion of assignment to Rifle Company or Special Training Battalion. The intermediate group, those recommended for assignment to the Special Training Battalion, was omitted from the analysis. Responses were transferred to IBM answer sheets so that an IBM graphic item count could be made. The percentage of correct attempts for the two groups was an index of item difficulty; the significance of differences in difficulty for the two groups, an index of item discrimination. The frequency of alternative responses on each item for the two groups was also obtained by IBM graphic item counts.

From these item analyses the best items were selected and the principles of good items hypothesized. On the basis of these generalizations, additional items for a revised form of the VC-1, X-1, were constructed.

Results

The R scores of the trainees in Rifle Companies, of the trainees recommended for Special Training, and of the trainees in Special Training Battalions were compared and the significance of differences between the means obtained. A similar analysis was made of the $R - \frac{1}{4}W$ scores. The only significant difference obtained for either scoring formula was that between those in Rifle Companies and those in Special Training Battalions. Table I summarizes these results.

*R = number of right answers; W = number of wrong answers.

TABLE I. Differences in Test Scores of Three Groups of Trainees

Group	N	Scored Number Right			Scored Right - $\frac{1}{4}$ Wrong		
		Mean	Critical Ratio		Mean	Critical Ratio	
			Rec. For Spec. Train.	Special Training		Rec. For Spec. Train.	Special Training
Rifle	62	48.72	1.11	3.49	44.82	0.99	3.31
Recomm. for Spec. Tr.	26	46.66		2.03	42.74		1.96
Special Training	43	42.12					

While those recommended for Special Training are closer to those in regular rifle companies than they are to those in Special Training Battalions with respect to VC-1 scores, the combination of the two groups, recommended for Special Training and in Special Training, was suggested on the basis of the original definitions of the "recommended" group. These men were to be those who would have been in the Special Training Battalion, had there been room enough for them. There was a marked amount of overlap in the distributions for the three groups, with standard deviations ranging from 8 to 11, the largest being for those in Special Training.

The correlations between R and R - $\frac{1}{4}$ W scores on VC-1 for Group A (Rifle Companies) and B (Recommended for, or, in Special Training Battalions) were practically perfect, being .99 for both groups. The regressions were obviously linear. On these grounds the scoring formula R - $\frac{1}{4}$ W was abandoned.

The reliability of VC-1, X-1 was estimated by both the Kuder-Richardson formula #21 and the split-half technique, correlating odd and even scores. TABLE II summarizes these results.

TABLE II. Estimates of the Reliability of VC-1, X-1

Group	N	Reliability		
		K - R #21	Split-half	Split-half raised by S-B *
Group A	62	.818	.824	.904
Group B	69	.875	.909	.952
Total	131	.864	.885	.939

While the K-R formula #21 gives an under-estimate of the reliability of a test, it is still highly probable that the true reliability is not as high as would be desired for a test of this type.

Estimates of the reliability of separate pages were also made by the K-R formula #21 in order to see whether or not a shorter test might be used as a preliminary screening device. The reliabilities were relatively low, however, ranging from .51 to .71. TABLE III gives the reliabilities for three pages of 20 items each.

* Spearman-Brown prophecy formula for test of doubled length.

TABLE III. Estimated Reliability of Separate Pages of VC-1, X-1

Group	Reliability by K-R #21					
	Page 2		Page 3		Page 4	
	N	r _{tt}	N	r _{tt}	N	r _{tt}
Group A	62	.623	62	.529	62	.514
Group B	69	.707	69	.678	69	.546
Total	131	.699	131	.666	131	.538

Page 2, which is the first page of the actual test, is the most reliable. Page 3 and 4 are progressively less reliable.

The reliabilities of DST-1 and DST-2 were also estimated by K-R #21. Since these tests had been considered for the same use as the VC-1, X-1, it was important to compare them on this characteristic. The DST-1 and DST-2 require less testing time, the time limits being 10 and 4 minutes respectively. Therefore, even equivalent reliability would favor the use of these tests in preference to the VC-1, ignoring any consideration of the relative validities of the tests.

TABLE IV. Reliabilities of DST-1 and DST-2 as estimated by K-R #21

Group	DST-1		DST-2	
	N	r _{tt}	N	r _{tt}
Rifle Companies	62	.702	62	.905
Recommended for Special Tng	26	.698	26	.898
In Special Training	43	.619	43	.939
Total	131	.708	131	.947

DST-1 was not sufficiently reliable, and was much less reliable than either DST-2 or VC-1, X-1. For the total group, DST-2 was more reliable than the VC-1, X-1. The reliabilities of DST-2 by K-R #21, the most conservative estimate of reliability, were considerably higher than the reliabilities of VC-1, X-1 estimated by the same method, and slightly higher than those estimated by the split-half technique.

Intercorrelations between the tests in the battery were obtained for Group A and B separately and combined. All of the intercorrelations were relatively low, none being as high as .50. The intercorrelations for Group A men in Rifle Companies, were lower than those for Group B or for the two groups together.

TABLE V.

Intercorrelations Between Tests given Regular Trainees (Group A)
and Those in Special Training or Recommended for Special Training (Group B).

TEST	OCT			2abc			DST-1			DST-2			MA-1		
	Group			Group			Group			Group			Group		
	A	B	A & B	AA	B	A&B	A	B	AA&B	A	B	A&B	A	B	A&B
VC-1,X-1	-.002	.232	.242	.360	.475	.460	.397	.424	.441	.263	.418	.427	-.019	.202	.200
N	62	65	127	62	65	127	62	69	131	62	67	129	56	47	103
GCT				.070	.271	.253	.115	.264	.268	.102	.265	.371	.208	.311	.329
N				75	68	143	62	65	127	62	63	125	69	51	120
2abc							.121	.356	.285	.059	.343	.327	.068	.299	.243
N							62	65	127	62	65	127	69	47	116
DST-1										.173	.381	.360	.078	.267	.240
N										62	67	129	56	47	103
DST-2													.281	.450	.470
N													56	46	102

The VC-1,X-1 correlated most highly with the 2abc, the correlation being .460 for the two groups together. The correlations of the VC-1, X-1 with the DST-1, and DST-2 were just slightly lower. All four of these tests were specially designed for use with men who were Grade V on the verbal intelligence tests. The Correlations of the Visual Classification Test with the GCT and MA-1 ranged from .00 to .24 for the three groups, the correlations for Group A being practically zero.

The number of cases for the intercorrelations varied from test to test, since all possible cases were used for each correlation. Table VI gives the means and sigmas on each test using the largest sample available in each group.

TABLE VI. Means and Sigmas on Six Tests for Groups
A, B, and A and B Combined.

TEST	Group A			Group B			A and B		
	N	Mean	Sigma	N	Mean	Sigma	N	Mean	Sigma
VC-1, X-1	62	48.8	8.1	69	43.9	10.3	131	46.2	9.6
GCT	75	60.5	6.9	68	52.3	8.9	143	56.6	8.9
2abc	75	62.2	5.0	68	59.8	9.7	143	61.1	7.7
DST-1	62	11.3	4.7	69	9.2	4.3	131	10.2	4.6
DST-2	62	21.3	7.2	67	11.2	8.8	129	16.0	9.5
MA-1	69	64.7	8.4	51	60.1	9.9	120	62.7	9.4

On every test except the DST-1 Group A was more homogeneous than Group B. This might have been partially responsible for the lower correlations observed for Group A. A limited range of ability was evident for both Groups.

and B, since the majority of men were Grade V on GCT and 2abc. However, since this is the range for which the VC-1, X-1 is intended, it was not desirable to estimate the intercorrelations for an extended range.

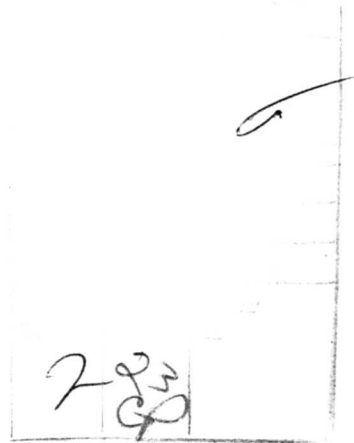
An item analysis of VC-1, X-1 was made to estimate item difficulty and validity on the basis of the external criterion of assignment to regular training versus assignment to special training. The percentage of correct attempts was used as an index of difficulty. If this value exceeded 90% for either group, the item was rejected, except where the value was much lower for the special training group indicating a high degree of discrimination for the item. Approximately half of the items were therefore eliminated as too easy for these groups.

Items answered correctly less than 10% of the time were discarded as too difficult. Very few items were in this category.

The significance of the difference between difficulty for the regular group and difficulty for the special training group was obtained for each item. Only three items had critical ratios of 2.5 or more, indicating that only a very few items discriminated between the two groups.

It was noted that the perception items tended to be too easy, and the paired comparison items non-differentiating. The best items were the abstraction items.

The frequencies of alternative responses on each item were examined for further guides in constructing items for the revised form of the test.



F P W
G