AD NUMBER

ADB951174

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; OCT 1952. Other requests shall be referred to Adjutant General's Office [Army], Washington, DC.

AUTHORITY

DoDD 5200.20

RESEARCH NOTES

(6)

ANALYSIS OF THE STUDIES OF CRITERIA AND PREDICTORS REPORTED IN DORCUS AND JONES "HANDBOOK OF EMPLOYEE SELECTION"

PJ-4901-02

(14) AGO-PRS-RESEARCH NOTE-52-70

(10) Daniel G. Severin

PERSONNEL RESEARCH SECTION
PR AND P BR, TAGO

ANALYSIS OF THE STUDIES OF CRITERIA AND PREDICTORS REPORTED IN DORCUS AND JONES "HANDBOOK OF EMPLOYEE SELECTION" AND OTHER LITERATURE

NATURE OF THE PROBLEM

BRIEF STATEMENT OF REPORT

This project was designed to investigate certain generalizations about correlations of criterion measures with various other criteria, the degree of relationship between ratings and personality tests, production records and personality tests, ratings and non-personality tests, production records and non-personality tests. One specific question was whether low test-criterion correlations were reported when the numbers of cases were small. Reporting bias was also involved in whether cross validated tests or tests that were not cross validated gave higher published correlations; and the same for "tailor-made" tests versus "ready-made" tests.

The question might be raised as to whether or not results of studies using various tests made by people with different ability, training, and experience could legitimately be grouped together. Obviously any findings from such heterogeneous material are suggestions rather than conclusions. Nevertheless, generalizations from these data may well prove useful as sources of hypotheses for further investigations.

BACKGROUND

Many validation studies have been done but seldom if ever have the correlations been classified and tabulated for the purpose of drawing generalizations about them. This project was designed to obtain such generalizations.

METHOD

CORRELATIONS USED

Most of the information was obtained from Dorcus and Jones, Handbook of Employee Selection(1). This is a reference book for both the layman and professional psychologist, consisting of 426 abstracts, of the literature on employee selection by means of psychological tests, from 1906 through 1948. Many of these studies are not reported in terms of correlations and so could not be used in the tabulations. Studies done by the Personnel Research Section, Personnel Research and Procedures Branch, The Adjutant General's Office, studies by psychologists in the Air Force, and other studies appearing in the literature too late to be included in Dorcus and Jones are also included in these tabulations.

It should be emphasized that the information obtained can only be considered representative of the studies reported; if there were a tendency to withhold certain kinds of correlations from the literature, then our analysis would not adequately represent all the correlations obtained.

CLASSIFICATION OF CORRELATIONS

The first classification dealt with correlations of various criteria with other criteria. A separate tabulation was made for each criterion with each different criterion. Nine different classifications were used, some of which were on-the-job ratings by superiors correlated with production records, on-the-job ratings by superiors with job proficiency tests, etc.

All the other classifications dealt with correlations of tests or predictors with criteria. Since negative correlations are as valuable for prediction purposes as positive ones the correlations have been tabulated in tables 2 through 5 regardless of sign. Personality tests were tabulated separately from non-personality tests (intelligence, performance, achievement, etc.) and each tabulated separately, by those using production records for criteria and those using ratings for criteria. Further breakdowns of the above were (1) by number of cases in the experiment, (2) whether the test had been cross validated, and (3) "ready-made" tests versus "tailor-made" tests.

RESULTS

Table 1 shows the correlations between various criterion measures. For example, column 1 shows three different correlations between on-the-job ratings by superiors and production records, with a median correlation of .46. The table gives the complete distribution; the range and median correlation for various criteria, as well as the number of correlations, which is an indication of the number of times psychologists have made studies of this type. If this table is a fair representation of the correlations found with such criteria, then certain generalizations can be drawn.

Training records and on-the-job ratings by superiors have a low degree of relationship, with a median correlation of only .11. In contrast with this, ratings by associates correlated quite highly with ratings by subordinates, the median correlation being .70. One fairly clear finding is that in general, criteria of training grades and on-the-job performance correlate low, with a median correlation of only .22. The median for all criteria correlated with other criteria in the table was only .28, which is evidence that extreme caution should be exercised in substituting one for the other.

Table 2 shows correlations of tests with on-the-job ratings and production records. The chief reason for making this tabulation was to get evidence concerning the hypothesis that personality tests predict ratings better than production records and non-personality tests predict production

Table [illegible] Correlations of various criterion measures

Corr.	ON-THE-JOB RATINGS BY SUPERIORS WITH: Pro. Rec.	[illegible] Assoc.	Profi-ciency Tests	Training by Assoc.	Training by Super.	[illegible] Records	ON-THE-JOB RATINGS BY ASSOCIATES WITH: [illegible] Ratings	Profi-ciency Tests	ON-THE-JOB RATINGS BY SUBORDINATES WITH: Pro. Records	Training Proficiency Tests With Pro. Records
.90-.94							.90-.94			
.85-.89							.85-.89			
.80-.84							.80-.84			
.75-.79							.75-.79			
.70-.74							.70-.74	1*		
.65-.69							.65-.69			
.60-.64				2			.60-.64			4
.55-.59							.55-.59			4
.50-.54							.50-.54			5
.45-.49	*	4		*			.45-.49			4
.40-.44		2					.40-.44			2
.35-.39		1*					.35-.39			3
.30-.34		1					.30-.34	1*	4*	3
.25-.29					2		.25-.29	2		3
.20-.24			3*	3	0*		.20-.24			9*
.15-.19					1		.15-.19		1	4
.10-.14						*	.10-.14			3
.05-.09							.05-.09			1
.00-.04							.00-.04			1
(-.05)-(-.01)							(-.05)-(-.01)			3
(-.10)-(-.06)							(-.10)-(-.06)			3
N	3	17	6	14	13		N	2	5	66
Med	.47	.13	.23	.47	.23		Med	.10	.32	.30	.24

*Median correlation.

Table 2. Personality and non-personality tests predicting job performance as judged by ratings and production records.

Correlation*	Personality Tests With			Correlation*	Non-Personality Tests With		
	Ratings	Prod. Records	Total		Ratings	Prod. Records	Total
.85-.89	1		1	.85-.89		1	1
.80-.84				.80-.84			
.75-.79				.75-.79		2	2
.70-.74	3		3	.70-.74	2	1	3
.65-.69	2		2	.65-.69	1	2	3
.60-.64	4		4	.60-.64	7	2	9
.55-.59	3		3	.55-.59	6	1	7
.50-.54	4		4	.50-.54	2	2	4
.45-.49	1	2	3	.45-.49	11	2	13
.40-.44	3	1	4	.40-.44	5	4	9
.35-.39	4		4	.35-.39	10**	5	15
.30-.34	7	5	12	.30-.34	10	7	17**
.25-.29	5	1	6	.25-.29	6	4**	10
.20-.24	6**		6	.20-.24	7	10	11
.15-.19	5	6	11**	.15-.19	5	5	10
.10-.14	8	8**	16	.10-.14	5	5	10
.05-.09	15	5	20	.05-.09	4	4	8
.00-.04	13	13	26	.00-.04	1	6	7
N	84	41	125	N	82	63	145
Med	.20	.11	.15	Med	.36	.26	.33

*Since negative correlations are as valuable for predictive purposes as positive correlations, the absolute values of the correlations have been tabulated in this and the following tables.

**Median correlation.

records better than ratings. This hypothesis has frequently been mention in psychological circles, presumably stemming from the belief that ratings depend more on personality qualities than do production records. Personality tests did predict ratings better than they predicted production records, but the hypothesis was not substantiated because the non-personality tests did not predict production records better than ratings. Ratings were more predictable by both personality and non-personality tests.

However, one may wish to consider an alternative statement of the hypothesis: that personality tests predicting ratings combined with non-personality tests predicting production records will give higher correlations than personality tests predicting production records combined with non-personality tests predicting ratings. To obtain an index of this, an average of the two medians (personality tests predicting ratings, .20; and non-personality tests predicting production records, .26) was computed and found to be .23. This was compared with the average of the two medians (personality tests predicting production records, .11; and non-personality tests predicting ratings, .36) which was found to be .24. This is again evidence that the original hypothesis was not substantiated. The average of the two medians was used rather than combining the two distributions and calculating the new median because the number of correlations with one combination (personality tests predicting ratings and non-personality tests predicting production records) would be too highly weighted with personality tests and the other combination (personality tests predicting production records and non-personality tests predicting ratings) would be too heavily weighted with non-personality tests. This is an important consideration because non-personality tests, in general, give higher correlations with job performance than personality tests and the difference found might be evidence of this factor rather than of the original hypothesis.

Another important finding was the great variation in correlations which ranged from .00 to .87. Other generalizations from this table seem to be that ratings are used as a measure of performance more than production records and that non-personality tests predict job performance better than personality tests (median correlation of .36 for non-personality tests as compared to .16 for personality tests).

The tabulation in Tables 3a and 3b was made primarily to see whether cross validation studies tend to report lower validities than non-cross validated ones. It was thought possible that cross validation studies might tend to be reported regardless of results, whereas pilot studies would probably not be reported if the results were disappointing. For all the correlations in Table 3b the test had either been cross validated before or the study reported was the first cross validation of the test. The number of studies used here is less than the other tables because if it could not be determined from the abstract whether or not the test had been cross validated, it was not used in the tabulation. The only place where a comparison could be made between tests cross validated and those not was in the case of the non-personality tests predicting on-the-job ratings. Here the median correlation

Table 3a. Comparison of cross validated and non-cross validated non-personality tests for predicting job ratings.

Correlations	Non-Personality Tests With Ratings	
	Cross Val.	Not Cross Val.
.85-.89		3
.80-.84		
.75-.79		
.70-.74	2	2
.65-.69		2
.60-.64	3	2
.55-.59	6	2
.50-.54	2	
.45-.49	11	1
.40-.44	3*	2
.35-.39	7	2*
.30-.34	7	3
.25-.29	4	1
.20-.24	3	2
.15-.19	2	3
.10-.14	2	2
.05-.09	3	2
.00-.04	1	
N	58	29
Med	.40	.38

*Median correlation.

Table 5b. Correlations of cross validated tests with job performance.

Correlations	Personality Tests With Job Ratings	Personality Tests With Prod. Rec.	Non-Personality Tests With Production Records
.85-.89			1
.80-.84			
.75-.79			2
.70-.74	5		1
.65-.69	3		1
.60-.64	2		1
.55-.59	1		1
.50-.54	1		2
.45-.49		2	
.40-.44	2	1	1
.35-.39	2		4
.30-.34	6	5	5*
.25-.29	5	1	1
.20-.24	4*		2
.15-.19	1	4	3
.10-.14	9	7	2
.05-.09	10	7	1
.00-.04	8	12	
N	57	57	29
Med	.20	.11	.34

*Median correlation.

was almost the same. So at least from this sample of correlations there is no evidence that only high non-cross validated validity correlations tend to be reported. Table 3b includes all studies where non-cross validated tests could not be found for comparison purposes.

Table 4 shows the increase or decrease of reported validity coefficients as the number of cases increased. The question is whether experimenters had a tendency to withhold low correlations and report only the high ones when small numbers of cases were involved. Correlations using a small number of cases can fluctuate by chance more than those with large numbers. But there was no consistent trend for the correlations to increase or decrease as the number of cases increased. The median correlation, of the non-personality tests with production records, increased as the number of cases increased; but for personality tests with production records, the correlation decreased as the number of cases increased. When ratings were used as the criterion there was no tendency for the correlations to increase or decrease as the number of cases increased. The same was also true when all tests and performance measures were totaled. Only about 4 percent of the experiments used more than 120 cases, and the median number of cases was about 50.

It is also interesting to note that the range in correlations had very little tendency to decrease as the number of cases in the experiment increased; one would expect some such decrease from statistical considerations. This would seem to be further evidence that the reporting of correlations was influenced in no consistent manner by the number of cases. This does not mean, of course, that studies with large numbers of cases are not more stable; it merely means that these tabulations show no evidence of writers reporting correlations which are spuriously high because low correlations were discarded whenever there was a small number of cases in the sample.

Table 5 is essentially a comparison of "ready-made" and "tailor-made" tests to see if there is a tendency for one type to have higher correlations than the other. The tests tabulated under personality tests do not include correlations with production records because no such personality tests were "tailor-made." The results suggest, however, that "tailor-made" personality tests are more satisfactory than the "ready-made" personality tests.

More information is included under the non-personality tests because here more "tailor-made" tests were found. Where non-personality tests were used to predict production records, the median correlation was about the same, .19 for "ready-made" and .20 for "tailor-made." Where on-the-job ratings were predicted the median correlations were .31 for both the "ready-made" tests and the "tailor-made" tests. The total for "ready-made" tests and "tailor-made" tests was computed for only non-personality tests because no "tailor-made" personality tests correlations with production records were found. Here the median correlation was higher for "ready-made" tests, .29 as compared to .22 for "tailor-made" tests. However, this difference was

Table 4. Validities of personality and non-personality tests for predicting job performance classified by number of cases in the experiment.

Corr.	Personality Tests With Ratings 0-40	Personality Tests With Ratings 41-80	Personality Tests With Ratings 81-	Personality Tests With Prod. Rec. 0-40	Personality Tests With Prod. Rec. 41-80	Personality Tests With Prod. Rec. 81-	Corr.	Non-Personality Tests With Ratings 0-40	Non-Personality Tests With Ratings 41-80	Non-Personality Tests With Ratings 81-	Non-Personality Tests With Prod. Rec. 0-40	Non-Personality Tests With Prod. Rec. 41-80	Non-Personality Tests With Prod. Rec. 81-	Total 0-40	Total 41-80	Total 81-
.90-.94							.90-.94									
.85-.89	1						.85-.89				1			2		
.80-.84							.80-.84									
.75-.79							.75-.79				2			2		
.70-.74		3					.70-.74	1	1		1			2	4	
.65-.69		2					.65-.69			1	1			2	2	1
.60-.64	1	2	1				.60-.64	2	3	2	1	1		4	[illegible]	3
.55-.59		1	2				.55-.59	2	2	2	1			3	[illegible]	4
.50-.54	2	1	1				.50-.54	1		1	2			[illegible]	1	2
.45-.49	1			1		1	.45-.49	4	5*	2	1	1		7	[illegible]	3
.40-.44	2	1				1	.40-.44		2	3	1	[illegible]		[illegible]	[illegible]	4
.35-.39	1	1	2				.35-.39	5	3	2*	4*	1		10	[illegible]	4
.30-.34	4	2*	1	1		3*	.30-.34	6*	1	3	4	2	1	15	[illegible]	6
.25-.29	4	1		1			.25-.29	5		1	1	2	1	11*	3	2
.20-.24	1	2	3				.20-.24	3	1	3	2	3*	*	6	11*	6*
.15-.19	3*		2	2	3	1	.15-.19	1	1	3	2	1	2	8	[illegible]	8
.10-.14	5	1	2*	1	6*	1	.10-.14	1	2	2	1	4		8	13	5
.05-.09	3	6	6	1*	3	1	.05-.09	3		1	1	3		8	12	8
.00-.04	7	2	4	7	6		.00-.04	1				6		15	14	4
N	35	25	24	14	16	8	N	55	21	20	27	32	4	111	9[illegible]	82
Med	.19	.31	.15	.05	.10	.31	Med	.32	.45	.35	.38	.21	.22	.29	.21	.25

*Median correlation.

Table 5. Validities of personality and non-personality tests for predicting job performance classified by ready-made and tailor-made tests.

Corr.	Personality Tests With: Job Ratings: Ready Made Tests	Personality Tests With: Job Ratings: Tailor Made Tests	Personality Tests With: Prod. Records: Ready Made Tests	Personality Tests With: Prod. Records: Tailor Made Tests	Corr.	Non-Personality Tests With: Job Ratings: Ready Made Tests	Non-Personality Tests With: Job Ratings: Tailor Made Tests	Non-Personality Tests With: Prod. Records: Ready Made Tests	Non-Personality Tests With: Prod. Records: Tailor Made Tests	Non-Personality Tests With: Total: Ready Made Tests	Non-Personality Tests With: Total: Tailor Made Tests
.85-.89					.85-.89			1		1	
.80-.84					.80-.84						
.75-.79					.75-.79			2		2	
.70-.74	1	2			.70-.74	2				2	
.65-.69		2			.65-.69			1		1	
.60-.64	2	1			.60-.64	4	1		1	4	2
.55-.59	1	3*			.55-.59	7		1		8	
.50-.54	2	1			.50-.54	1		2		3	
.45-.49	1		2		.45-.49	12	1	1	1	13	2
.40-.44	3		1		.40-.44	3		1	3	4	3
.35-.39	4	1			.35-.39	12	1	4	1	16	2
.30-.34	10		6		.30-.34	10*	3*	4	2	14	5
.25-.29	8		1		.25-.29	9	1	1	2	10*	3
.20-.24	11				.20-.24	6	2	2	6*	8	10*
.15-.19	11		4		.15-.19	10	1	8*	1	18	2
.10-.14	17*		6		.10-.14	9		7	4	16	4
.05-.09	26		4*		.05-.09	7		5	3	12	3
.00-.04	23		17		.00-.04	4		3	7	7	7
N	120	10	41		N	96	10	42	32	139	43
Med	.13	.60	.09		Med	.31	.31	.19	.20	.29	.22

*Median correlation.

largely if not entirely due to the large proportion of "ready-made" tests predicting ratings rather than production records. Such is the case because, in general, higher correlations are obtained when ratings are predicted compared to production records.

SUMMARY

The data presented herein seem to warrant the following generalizations:

1. Training records correlate low (median about .20) with on-the-job performance and should not replace on-the-job performance without testing the assumption that they are equivalent.

2. In general, care should be used in substituting one criterion for the other as the median correlation of job performance criteria with various other criteria is only about .30. These correlations range from .00 to .87, which would seem to be strong evidence that at times the research worker can substitute one criterion for the other after first determining that they were reasonably equivalent.

3. The hypotheses that personality tests predict ratings better than production records, and non-personality tests predict production records better than ratings, were not substantiated.

4. Non-personality tests give higher predictions for production records and ratings than personality tests, a median of .33 as compared to .15.

5. No evidence was found that writers tend to disregard correlations that are low when there are small numbers of cases in the sample.

6. Only about 4 percent of the experimenters validated their tests with 120 or more cases. The median number of cases in the validation studies was about 50.

7. In the studies reported here, "ready-made" non-personality tests appear to be no better or poorer for predicting job performance than "tailor-made" tests. However, the evidence (from the same source as above) suggests that "tailor-made" personality tests are more satisfactory for predicting job performance than "ready-made" personality tests. Presumably, the specificity of the job under consideration would make a lot of difference, but we have no information on that factor.

CONCLUSIONS

1. Job ratings by associates agree better with subordinates than with superiors' ratings.

2. Personality tests seem to agree better with job ratings than with production records.

3. Non-personality tests agree a bit better with job ratings than with production records.

4. "Tailor-made" personality tests are probably preferable to the "ready-made" kind, but this is not necessarily true of non-personality tests.

5. Criteria cannot be substituted for one another without first knowing their degree of equivalence.

PERSONNEL IN CHARGE

Program Coordinator: Dr. Richard M. Gaylord
Acting Project Director: Dr. Daryl G. Severin
Statistical Adviser: Mr. Cecil Johnson
Preparation of Report: Dr. Daryl G. Severin

COLLECTION OF DATA: September 1951

PREPARATION OF REPORT: 25 October 1951

REFERENCE

1. DARNELL, R. M. and JONES, M.H. Handbook of employee selection. New York: McGraw-Hill Book Company, Inc., 1950.