MASTER FORMAT

RESEARCH MEMORANDUM 71-3

ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK: AIRFIELD LAYOUT

DDC FILE COPY

DDC
RECEIVED
DEC 20 1979
A

U. S. Army

Behavior and Systems Research Laboratory

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

April 1971

79 12 18 331

Army Project Number
2O662106A723

Officer Prediction d-82

Research Memorandum 71-3

ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK: AIRFIELD LAYOUT

Edward M. Sait

Louis P. Willemin, Task Leader

ECI RM-71-3

Accession for	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

Submitted by:
William H. Helme, Chief
Behavioral Evaluation
Research Division

Approved by:
J. E. Uhlaner, Director
U. S. Army Behavior and
Systems Research Laboratory

April 1971

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Behavior and Systems Research Laboratory.

402 777

CONTENTS

ANALYSIS OF OFFICER PERFORMANCE ON AN EXPERIMENTAL TASK: AIRFIELD LAYOUT

The Airfield Layout Task is one of fifteen situational performance tests developed and administered within the OFFICER PREDICTION Work Unit.

The Officer Prediction Work Unit is a large-scale longitudinal research project initiated in response to recommendations by the Army Scientific Advisory Panel (ASAP) and by the Deputy Chief of Staff for Personnel (DCSPER). The former indicated a need for additional research on the performance and selection of combat officers and suggested that dimensions of such performance might be defined by means of performance exercises within combat simulation. DCSPER, in view of the increasing complexity of military technology, was interested in determining the feasibility of differential prediction of performance in broad areas of possible officer specialization. The research design incorporates both sets of requirements.

The research is concerned with three broad areas of officer activities--combat, administrative, and technical. Experimental predictor tests relevant to these areas were administered to 6000 officers at entry on active duty in 1958 and 1959, and a revised battery to 4000 at entry on active duty in 1961-1964. Fifteen to 30 months later, a subsample of 900 of the latter group, six at a time, were assigned to the Officer Evaluation Center, Fort McClellan, Alabama, established to obtain criterion data on officer performance. There, in a simulated Military Assistance Advisory Group (MAAG) setting, a scenario unfolded which eventuated in invasion and guerrilla warfare. Over a period of three days, the six officers received a series of assignments, first administrative and technical, then combat. Performance was recorded and rated out of sight of the examinee by cadre who played the parts of MAAG, host nation, and aggressor personnel. Work products were retained for later scoring. The performance records and work products, after analysis to define underlying dimensions, served as criteria for the predictor tests.

The Airfield Layout Task was one of the five in the technical area. It was administered on the second day. The day began at approximately 3 A.M. with briefings on outbreak of hostilities and on road-damage and radiation surveys to be made. After the briefings, three of the examinees were assigned the Airfield Layout Task to be completed while they waited for their reconnaissance teams. The other three examinees were given the Airfield Layout assignment at approximately 7 P.M. The difference in time, greater than that between administrations of any of the other fourteen tests, permitted staggered start of other tests for efficient use of OEC staff.

According to the scenario, a "hasty" airfield was required in an area to the northeast. The examinee was asked to prepare a report summarizing, for each of three proposed sites, information relevant to evaluation

of the site on tactical, operational, and engineering grounds. The examinee was given manuals which treat the important considerations, a map of the area with slope and soil overlays, and a report of observations made on an air reconnaissance of the area. He was also asked to compute the length of the runway that would be required to handle airplanes of a specified type. Other manuals given him contained the procedural, aircraft, and additional geographic data necessary for the computation. He had one hour for both parts of the assignment. Scoring of his report was done on a 26-item Airfield Layout Checklist. A single credit was given for each type of evaluative consideration adequately presented in the report for one or more of the three sites. One additional evaluation was made: a rating of how well he appeared to understand the mission after it was explained to him.

The test was intended to measure ability 1) to obtain necessary evaluative criteria and procedures from references; 2) to obtain from various sources--manuals, maps, and reports--the data required to make the evaluations or follow the procedures; and 3) to perform a prescribed sequence of simple arithmetic operations on data obtained from references.

Partly because the test was one of the last to be developed and introduced, certain changes appeared necessary in instructions and in scoring during the post-shakedown period at the OEC. When the test was first given, beginning with Group 20, five sites were to be evaluated, too large a job for the time allowed. Beginning with Group 39, the requirement was changed to three, and a tabular format was mentioned in the briefing given the examinee as an acceptable way to present findings. Then, because many examinees were giving evaluations without supporting facts (e.g., a site might be evaluated as only fair with respect to psychological hazards without indication that the reason was a church steeple near the glide path), specificity in reporting was called for in all briefings starting with Group 47. With Group 53, this instruction was emphasized by an example; and mention of a tabular format, sometimes associated with inadequate specificity, was omitted. Furthermore, just prior to the testing of this group, additional instructions were given the examiners to reduce misunderstandings of and inconsistencies in application of scoring standards.

OBJECTIVES

The main objectives of the analysis were:

1. To evaluate checklist content at item level.

2. To evaluate the effect of and, if desirable and possible, to correct for various potentially disturbing variables. Conspicious

among these are time of day, examiner, sequence in which the examinee undertook his work (site evaluation or runway length undertaken first), and calendar periods associated with altered instruction or scoring. In addition, the scoring procedure, in which no additional credit was given for application of an evaluative criterion to more than one site, seemed likely to have reduced the comparability of scores.

3. To estimate reliability of the checklist.

4. To evaluate the extent to which scores were affected by background, as represented by membership in the Corps of Engineers vs other branches of the Army.

5. To identify dimensions (factors) in the scoring record.

6. To provide scores for major parts of the scoring record, for the dimensions, and for the test as a whole. These scores are to be correlated with scores from the other fourteen tasks to indicate the extent to which each is specific to this task, common to the tests of the technical area, and general across all three areas. From these scores and those of the other tasks, criterion scores will then be derived to validate the experimental predictor tests.

METHOD

SAMPLES

The total sample consists of all cases (786) for whom scoring records were obtained after trial runs were completed. For several analyses, only those records from Group 53 on, when procedures became stabilized, were used (N = 617). For special purposes, other subsamples were used.

VARIABLES

The basic scores are derived from the 26 items of the Airfield Layout Checklist. The first 10 items, on computation of runway length, provided a runway total; the remaining 16, a site-report total. The checklist also provided a supplementary measure, a rating on a five-point scale of the examinee's understanding of his mission, indicated by his questions during the briefing period and his ability, when requested, to recapitulate the mission.

Certain supplementary scores were obtained by inspection of the examinee's papers. Two of these are measures of achievement: extent of completion of runway computation (coded 0, did not begin; 1, partial completion; and 2, completion) and number of sites covered in the report.

In order to evaluate the effect of order of work (sites first or runway first) on scores obtained, an inferred order-of-work score was derived for 359 examinees who completed only one of the two parts of the assignment. Two other scores relate to understanding of instructions: factualness of the site report and avoidance of a personal selection of a site. Evaluations of factualness of site report were coded 1, predominantly evaluative; 2, intermediate; 3, predominantly factual or with factual support for nearly all evaluations; and 4, factual but bearing of facts on evaluation not indicated. The last code, a reversal of the prior progression from undesirable to desirable was so infrequent as to have little effect on the mean and standard deviation of the factualness score. Evaluations of avoidance of personal selection were coded 1, supported one site at expense of adequate presentation of data on the others; 2, selected one site but gave approximately equal information on the others; and 3, no selection made.

Other variables represent the aforementioned conditions of administration which could give rise to error: time of day, examiner, and calendar period (four successive periods were established, with changes in test requirements at start of second and third).

The variable, branch school attended, was used in examining sensitivity of scores to differences in background represented by Corps of Engineers vs other branches.

ANALYSIS

Item p-values and intercorrelations were obtained to detect any items with such extreme p values or lack of positive intercorrelation as to warrant elimination. In addition, p values for the subsample that had had the engineering basic branch course were obtained and compared with the p values of the remaining examinees. Any item on which this presumbly more knowledgeable and abler subsample did not do as well or better was to be examined for possible elimination or rekeying.

Various approaches were used to study the effects of possible sources of error. The frequencies of scores representing departure from instructions(and possible misunderstanding of them) or other potential sources of error were determined. Correlation between scores representing time of day, departure from instructions, and order of work (site or runway first) with site, runway, and total scores were obtained. Means and standard deviations on the latter scores were compared for subsamples representing different values of potentially disturbing variables. An analysis of variance was conducted for time of day, examiner, and calendar period.

Relationships were examined between scores and the background variable, branch school attended.

A principal-component factor analysis of the 26 checklist items was undertaken. So that speededness and sequence-of-work effects would not prevent emergence of factors across site and runway subtests or otherwise distort results, the analysis was based on the 294 examinees who completed both parts of the test. Tetrachoric correlation coefficients were used. Results were rotated by the varimax procedure. Factor scores were established by selection of variables with high loadings on a given factor, low loadings on others.

For use in across-test and other analyses, a total-score formula was established.

RESULTS AND DISCUSSION

ITEM ANALYSIS

All items were judged acceptable with respect to p values. These ranged from .08 to .88 for the entire sample, with a mean of .49 on the site-report items and .35 on the runway-computation items. These values were affected by the failure of many individuals to complete the test or even to begin work on the runway computation. For the 294 who completed both parts, the range was .10 to .94, and the respective means were .50 and .58.

Items were also evaluated by comparing the p values for 59 examinees who had attended the Corps of Engineers basic course, and who by aptitude, training, and experience should have an advantage on the test with the p values for the remaining examinees. On 20 of the 26 items, the engineer subsample had higher p values, on one the same, and on five lower. However, none of the latter five differences were statistically significant. Moreover, all pertained to the site report, on which the engineers' advantage was likely to be less than on the runway computation, and all five items appeared sound in content. Therefore, no items were eliminated on the basis of this comparison.

Tetrachoric item intercorrelations were examined in the sample of 294 examinees who completed the test. (Though intercorrelations within this subsample may be reduced through restriction of range, their relative magnitudes should better represent inherent relationships among items than would intercorrelations within the entire sample, where the effects of speededness would be more pronounced, greatly inflating some intercorrelations and reducing others across subtests.) In the subsample, there was a fairly large number of negative intercorrelations--21% of the coefficients within subtest (site-report or runway computation) were negative and 43% across. Nearly a third of the negative values reached the .05 level of significance. The negative intercorrelations may be due largely to heterogeneity of content, with competition of content areas for the examinee's time, rather than to item defects. For each item, the highest coefficient was at least +.30, and the average of coefficients

positive. Consequently, no items were eliminated on the basis of item intercorrelations.

MEANS, STANDARD DEVIATIONS, and INTERCORRELATIONS OF MAJOR VARIABLES

Means, standard deviations, and intercorrelations of major variables are shown in Table 1. Besides basic scores, the table includes certain supplementary scores used in study of possible sources of error in the basic scores. The negligible correlation between the site-report and runway scores is attributable to the examinee's freedom to allocate his limited time as he sees fit between the two parts of the task. The negative correlation of site-report and of total score with undertaking the site report first, rather than the runway first, is discussed below under "Error associated with scoring procedure and order of work."

SOURCES OF ERROR

Circumstances of Administration. The effects of three variables, which, both on the basis of observation at the OEC and of variations in mean scores, seemed likely to have affected the test scores, were treated by analysis of variance. The variables are time of day (morning vs evening), calendar period (earlier cases, Groups 53-100 vs later cases, Groups 101-159), and examiner (six examiners who conducted most of the Airfield Layout testing over these two periods). In the design of the analysis, examiners were nested within time of day, because each examiner was regularly assigned either to morning (3 examiners) or to evening (3 examiners). Differences in cell size were handled by use of the harmonic mean. A fixed-effects model was employed for determining F-ratios, because interest lay in the effects of the specific conditions at the OEC rather than in generalizing beyond them. Analyses were performed separately for site report, runway computation, and total score.

The results (Table 2) show significant main effects for each of the three administrative variables on site-report scores but not on runway scores. The result for time of day is consistent with correlation coefficients shown in Table 1. The tendency to better performance in the evening might be expected in view of the frequent finding that human efficiency is low during early morning hours. Examiners may affect scores in a variety of ways--through degree of clarity in giving instructions, through ability to motivate, through relative emphasis on the two parts of the task, and through their additional role of scorer. Calendar period provided the most significant effect, with higher scores in the earlier period. Significance carried over to the total score. Interpretation here is uncertain. The effect may represent systematic error--for example, elimination of unauthorized hints during instruction or a tightening of scoring standards--or, on the other hand, a systematic change in the examinees scheduled to come to the OEC.

For the following reasons, no corrections in score were made for the three variables whose effects were tested. The significant effects were on one part of the test, the site report, only. The magnitude of score

Table 1

INTERCORRELATIONS OF AIRFIELD LAYOUT VARIABLES

(GROUPS 53 - 159, N = 615[a])

Variable	Mean	S.D.	Intercorrelations									
Time: AM vs PM	-	-	Time									
Understanding Mission	3.21	.72	.00	U.M.								
Order: Site 1st vs RW 1st[a]	1.77	.42	.00	-10	Order							
Factualness of S.R.	2.88	.49	.14	.03	.15	F						
Avoiding Personal Selection	2.80	.53	.20	.07	.23	.36	A.P.S.					
No. of Sites Completed	2.59	.72	-.01	.00	.53	.13	.27	N.S.C.				
Runway Completion	1.26	.83	-.02	.12	-.83	-.12	-.08	.17	R.C.			
Site Report Score	7.66	2.89	.10	.20	.18	.33	.34	.33	.01	Site		
Runway Score	3.64	2.77	.00	.15	-.75	-.09	-.05	.16	.87	.04	RW	
Total: Site Plus RW	11.31	4.10	.07	.24	-.37	.17	.21	.34	.60	.74	.71	

[a] Order of work (site report first vs. runway computation first) could be determined for only 359 of the 615 examinees. Statistics for this variable are based on the 359. A score of 2 represents sites first; 1, runway first.

Table 2

ANALYSIS OF VARIANCE OF CHECKLIST SCORES
BY TIME OF DAY, EXAMINER (NESTED),
AND CALENDAR PERIOD[a]

		Site Report		Runway		Total	
Source	df	MS	F	MS	F	MS	F
Time (T)	1	39.98	4.88*	.55	.07	33.46	2.09
Examiners (E)	4	23.60	2.88*	10.21	1.38	22.76	1.42
Period (P)	1	79.91	9.75**	.42	.06	79.85	4.98*
T x P	1	2.77	.34	.57	.08	3.89	.24
E x P	4	2.64	.32	7.43	1.00	.28	.02
Within cells	574	8 20		7.40		16.01	
Total	585						

*p = .05
**p = .01

[a] Time of day is morning vs. evening; calendar periods are (1) earlier cases after stabilization of procedures (Groups 53-100) and (2) later cases (Groups 101-159).

difference associated with (but not necessarily entirely attributable to) evening vs morning administration was not large, about one fifth of a total-sample standard deviation. Examiner deviations from the general mean ranged up to about one quarter of a standard deviation, but, due to smaller Ns, with a still larger possible chance component in the deviations. As indicated above, the calendar--period effect may have been due to systematic change in sampling and not to the error of administrative change. Further study of these areas of error and possibly resulting score correction appear warranted only if particularly refined scores are required.

Understanding of Instructions. Some examinees apparently misunderstood the purpose of the site report as selection and justification of one site over others, rather than presentation of impartial data useful in later selection. Also, some examinees, presumably because of misunderstanding, evaluated each site in such terms as "good" or "poor" with respect to various criteria, without presenting the factual bases for the judgments, as was intended. Correlation coefficients of the checklist site-report score with scores representing avoidance of these tendencies, as judged by inspection of the site reports, were .34 and .33, respectively; the multiple R with these two scores was .41.

Occasional inadequacies in instruction undoubtedly contributed to these tendencies. Also, perhaps, prescribed wording of instructions might have been improved. However, whether and to what extent a correction in site-report scores for reduction due to these tendencies would be justified is uncertain. Misunderstanding of the task must in some cases, and at least to some extent, be attributable to the examinee, and it may be appropriate that the score to some extent be affected by and thus measure abiility to attend to, comprehend, and remember the instructions. Second, part of the lower level of scoring associated with these tendencies may be due to lower status on abilities employed both in understanding and executing instructions. Third, some examinees may have selected and supported a single site, or resorted to unsubstantiated evaluative adjectives, as a shortcut when pressed for time, even though they knew they were not fulfilling test requirements. (This possibility is suggested by correlation coefficients of only .07 and .03, respectively, between the avoidance of the two tendencies and understanding the mission, although the latter had significant correlation (.20) with the site-report score.)

Error Associated With Scoring Procedure and Order of Work. Fifteen of the sixteen site-report scoring items represent factors that should be considered in the evaluation of each of the three assigned sites. In the scoring of these items at the OEC, unit credit was given (a "Yes" checked) whenever the evaluative factor was adequately treated in the report on any one of the three assigned sites. No additional credit was given for its satisfactory application to one or both of the other sites, nor was credit deducted for absence of such application. As a result, report on a second and on the third site tended to produce relatively small increments in score. Although there may be some tendency to greater thoroughness among

examinees who complete fewer sites, the data of Figure 1 are illustrative. The mean Site Report scores of examinees who completed one, two, and three sites were, respectively, 5.7, 7.0, and 8.0. (The slight fall-off shown in the figure for three sites completed as compared with the third started but not completed may represent relatively frequent sacrifice of thoroughness for speed in the group that completed all three assigned sites.) Over one-third of the examinees completed fewer than three sites. These examinees, then, received liberal scoring.

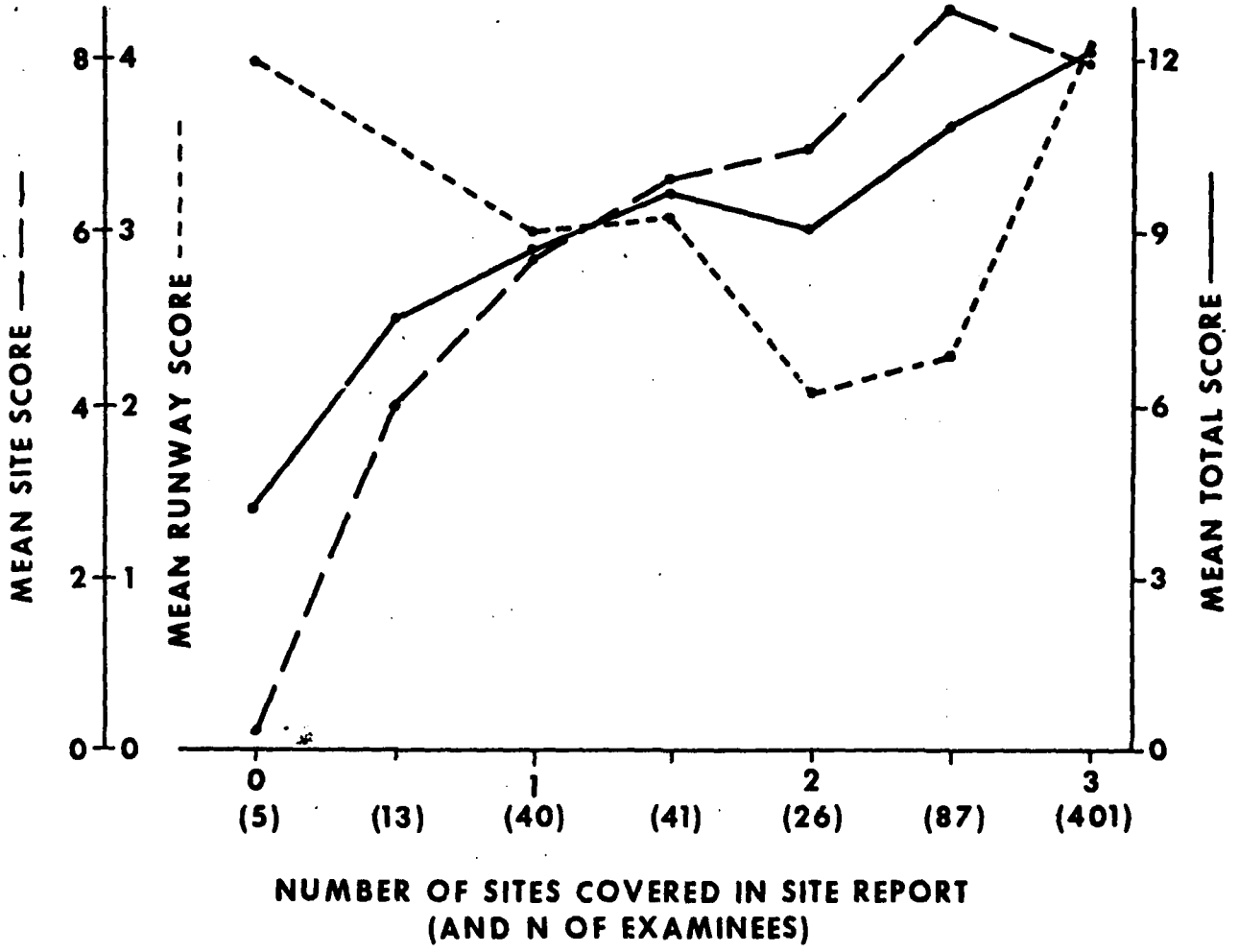

Figure 1. Relationships between Number of Sites Covered in Site Report and Score.

Of particular concern, the scoring procedure seems often to have given substantial advantage to those examinees who began work with the runway computation, which was for most much the less time-consuming part of the assignment. These examinees almost always had time left to prepare at least a partial site report, which, under the scoring procedure, might receive nearly as much scoring credit as a complete report. But examinees who began with the site report were often still working on it at the end of the allowed time and thus received a zero runway score. Less than one percent failed to begin the site report; twenty-five percent failed to begin the runway computation. For 359 examinees whose order of work could be determined from inspection of work products, starting with the site report rather than with the runway had a negative point-biserial correlation of -.75 (Table 1) with runway score but only a +.18 correlation with site score. The overall effect is indicated by a -.37 with total score. The mean total score of the 84 examinees identified as having undertaken the runway first was 12.2, 37% higher than the mean, 8.9 of the 275 identified as having undertaken the site report first. In some instances, poor quantitative ability may have led to decision to leave the runway to the last. However, it is considered doubtful that any large part of these effects of order of work can be attributed to poorer quantitative ability of those who chose to begin with the site report.

Rescoring to give due credit to second and third sites should give more equitable scores. However, an excessive amount of time would be required for this. A rough correction was undertaken as described under "Final Scores," by incorporating in the report score an increment proportionate to the number of sites undertaken. This was intended to bring the site-report score into approximate alignment with the expected result of rescoring.

RELIABILITY

In the hour given to complete the Airfield Layout Task, only 60% of the examinees (Groups 53 through 159) completed the site report and 51% the runway computation. Therefore, the task can be considered a speeded test to which internal-consistency measures of reliability are of doubtful applicability. However, because these are the only reliability measures available, consideration is given below to interpreting and deriving from them an evaluation of test reliability.

In the case of the runway computation, most of the scoring items represent steps which were carried out in a uniform or fairly uniform sequence. Speededness in such a situation will inflate the correlation between odd-even or matched-content halves used to compute a split-half reliability coefficient. It will also tend to inflate item intercorrelations, thereby test variance in relation to sum of item variances, and by this means the Kuder-Richardson 20 reliability coefficient. Failure of many examinees to begin the runway computation added to the inflation of internal-consistency coefficients.

In the case of the site report, however, time pressure may have had little inflationary effect on internal-consistency reliability measures. There was no sequence in which scoring points had to be made. Though most examinees reported on sites in the order in which they were presented, this procedure did not necessarily introduce a sequence among items, for most scoring points could be made about equally easily in reporting on one site as in reporting on another. A general tendency to take up in uniform order the three main considerations--engineering, tactical, and operational--might tend to introduce a sequence. But the effect of any such tendency could not be pronounced, for most examinees completed a report on at least one site and so had opportunity to treat all three considerations. Thus, in absence of an item sequence with later items attempted less often, and in absence of an appreciable number of examinees who failed to start this part of the test, time pressure seems unlikely to have had a systematic inflationary effect on internal-consistency reliability. There may instead be an opposite effect, through attenuation of item intercorrelations because of an effectively chance determination of items missed under time pressure.

It thus appears that an internal-consistency coefficient gives a fairly conservative estimate of site-report reliability but an inflated one of runway-computation reliability. Obtained Kuder-Richardson 20 coefficients were .65 and .84, respectively. Under the assumptions that the first coefficient is acceptable and that runway and site items are of the same level of internal-consistency reliability, the Spearman--Brown formula was applied to the site coefficient to obtain a reliability estimate for the runway computation (for which there were 10 items vs 16 for site). The result was a coefficient of .54.

Another approach to internal consistency measurement of reliability is to derive coefficients from the relatively small sample that completed the test. This would so far as possible remove the inflationary effect of speededness. In particular, it would remove the inflationary effect of failure to attempt one part of the test, as was frequent in the case of the runway computation. However, a tendency toward underestimating is introduced through restriction of ability range, and correction for this restriction through use of the correction formula in which the total-sample variance is entered would be inappropriate. (The greater magnitude of this variance over the restricted-sample variance cannot be attributed entirely to greater range of ability at the low end, but is due in part to failure to attempt the problem for other reasons than inability.) The coefficients obtained in this subsample, and considered for the reasons above to be underestimates were .60 and .48, for site and runway respectively. These values are close to, and hence support, the previous estimates of .65 and .54.

Reliability coefficients of .65 and .54 for the two tasks within a test are rather low. The low level of these estimates may be attributed

in part to the relatively small number of checklist scoring items (26) as compared with the number in other OEC situational tests and in part to the somewhat heterogeneous content of the runway and site tasks. Consistency of scores across parallel forms of the test might be higher. However, high reliability coefficients should not be expected if examiner, time of day, and other causes of impaired comparability which occurred in the actual testing were allowed to vary randomly across the two administrations.

EFFECT OF ARMY BRANCH

The extent to which score variance could be accounted for by differences in prior training and experience was of interest. Strong dependence on such differences would indicate a need to correct scores for experience before they used as criteria for development of instruments for prediction of performance or for selection, guidance, and assignment.

One available item of background information which should reflect sensitivity of scores to differences in training and experience is branch of the Army membership. Of the branches represented at the OEC, the Corps of Engineers has particular relevance to the Airfield Layout Task: a substantial proportion of the officers assigned to this branch have a college degree in civil or, less frequently, in other fields of engineering; airfield construction had been covered in the basic branch course; and a few of the examinees had done some subsequent work in this area. Table 3 presents score means and dispersions for the Engineers and for the other branches represented, grouped, first, into two other technically oriented branches, then into combat, with Air Defense separated out as atypical, and finally into the primarily administrative. On all possible comparisons, the Engineers were superior, except on the site-report score in which their performance was even with that of Finance examinees. In terms of the total-sample standard deviation, the mean of the Engineers on the runway task was about three-quarters of a standard deviation above the other examinees; their mean on the total score, about two-thirds of a standard deviation above. Three quarters of the Engineers were above the average of the other examinees on these scores. The differences between Engineers and each of the other branches were individually significant at about the 5% level or better. Variance of the ten branch means, equally weighted, for the three scores were .72, .30, and 1.28, or approximately 9%, 4%, and 8% of the total variances. Thus, the test appears quite sensitive to specifically relevant training and experience such as represented by the Corps of Engineers, but fairly independent of training and experience as represented by the gamut of branches represented among examinees.

The first observation is subject to qualification. The indicated superiority of the Engineers is not necessarily attributable entirely to

Table 3

SCORES BY ARMY BRANCH

(Groups 53 on)

Branch School	N	Runway M	Runway SD	Site Report M	Site Report SD	Total M	Total SD
Engineers	59	5.3	2.6	8.2	2.9	13.5	3.8
Ordinance	99	4.0	2.7	7.9	2.7	11.9	3.7
Signal	54	3.8	2.8	6.6	3.0	10.4	3.9
Mean		3.9	2.7	7.3	2.9	11.1	3.8
Armor	72	3.0	2.5	7.4	2.2	10.4	3.5
Artillery	75	4.0	2.9	7.9	3.2	11.9	4.5
Infantry	93	3.2	2.8	7.9	2.9	11.1	4.1
Mean		3.4	2.7	7.7	2.8	11.1	4.0
Air Defense	53	4.1	2.7	7.2	3.1	11.3	4.8
Quartermaster	53	2.7	2.6	7.8	2.9	10.5	4.0
Adjutant General	21	2.2	2.1	6.7	2.3	9.0	2.8
Finance	36	2.9	2.5	8.2	3.0	11.1	4.0
Mean		2.6	2.4	7.6	2.7	10.2	3.6
Mean, other branches than Engineers		3.3	2.6	7.5	2.8	10.8	3.9
All cases combined	615	3.6	2.8	7.7	2.9	11.3	4.1

training and experience. Selective factors are likely to draw into the Corps of Engineers individuals with aptitudes appropriate both to that branch and to the test. The second observation should be interpreted with caution. Scores were doubtless influenced also by many background variables other than and not closely correlated with branch, and the sum of these influences may be substantial.

FACTOR ANALYSIS

The 26 items were factor analyzed as described. The six-factor solution was accepted, accounting for 48.8% of total variance as compared with 45.3 and 52.0 with the five- and seven-factor solutions. The six factors are listed below. For each factor, the higher loading items which contribute to the definition and serve also as a factor score are designated by number and listed in order of the magnitude of the loading, which is given in parentheses.

1. Operational hazards
   - 21. Stream or lake listed as psychological hazard (.84).
   - 22. Town or high building listed as psychological hazard (.77).
   - 18. Possible ground fog noted (.61).

2. Altitude and grade correction.

Interpretation is somewhat doubtful. A tendency to thoroughness may lead to credits on items 4 and 8, mention of altitude and grade corrections when considered not applicable, as well as to identification of figures arrived at. However, grade correction, though scored 0, could be justified and may presumably have been made by examinees who are meticulous.

   - 4. Indication that altitude correction is not applicable (.74).
   - 8. Grade correction given as zero or not applicable (.65).
   - 2. Correct identification of all final figures (.55).

3. Engineering and tactical considerations.
   - 13. Source of water noted (.68).
   - 17. Terrain evaluated with respect to protection against attack (.58).
   - 16. Drainage problem mentioned (.52).
   - 15. Source of fill (e.g., gravel) noted (.48).
   - 14. Access roads noted (.44).

4. Correct safety factor and consideration of surface softness in runway computation.

6. Correct safety factor (.93).

7. Softness of surface considered (.68).

10. Final runway length within acceptable limits (.68). This item also has high loading (.61) on Factor 2.

5. Use or removal of buildings and vegatation.

12. Use or removal of buildings (.58).

23. Need to clear vegetation noted (.54).

24. Availability of vegetation for camouflage (.31).

6. Geographic considerations affecting site suitability.

20. Prevailing wind considered (.53).

26. Degree of slope noted (.47).

11. Nature of soil noted (.47).

FINAL SCORES

The items of the two parts, site report and runway computation, were judged adequate with respect to p values and intercorrelations and appeared to give a fairly balanced coverage within and across these parts. A refined differential weighting of these items for a final total score was not undertaken, a process which would appear justified only if an independent, more nearly ultimate criterion were available for the purpose. However, in the cases of the site-report and total scores, it seemed desirable to provide some correction, short of rescoring, for error variance resulting from the scoring procedure which had given no additional credit for applying an evaluative consideration to more than one of the three assigned sites, thus favoring thoroughness in the tradeoff between completion of assignment and thoroughness, and which favored those who began work with the generally more quickly completed runway part of the assignment.

The major means adopted for correction was to add to the site report score a multiple of the score proportional to the number of sites covered in the report. A second means, employed for the total score, was to reduce the relative contribution of the runway score, which was the one most affected by order in which work was undertaken, sites first or runway first. The extent of both corrections was based on subjective judgement. The formula for the corrected site-report score is:

$$S_c = (n + 5)S$$

where S represents the uncorrected site-report score and n is given values as follows:

n	Sites completed
1	One site started
2	One site completed
3	Second site started
4	Second site completed
5	Third site started
6	Third site completed

The constant multiplier 5 reduces the force of the number-of-sites multipliers to allow for somewhat greater average thoroughness in incomplete reports and for the possibility that certain scoring points may be more easily obtained on the second site undertaken or possibly on the third.

The modified total score was obtained as follows:

$$T = S_c + 5R,$$

where S_c is the corrected site-report score and R the runway score. The multiplier 5, in conjunction with an average multiplier of approximately 10 for the raw site-report score, represents a reduction in relative weighting of the runway score over weightings represented by simple addition of original site-report and runway scores in either raw or standard score form.

Items providing factor scores are listed in the preceding section on pages 15 and 16.

SUMMARY AND CONCLUSIONS

The Airfield Layout Test was one of five situational performance tests administered at the Officer Evaluation Center to measure ability in technical assignments, as distinguished from administration and combat assignments. The examinee was required to report on the adequacy--with respect to tactical, operational, and engineering considerations--of three proposed sites for a hasty airfield and to calculate the length of runway needed.

Data from 786 examinees, and subsamples of these, were analyzed to evaluate the test as a partial criterion of technical performance. Analysis led to the following results and conclusions:

Item Statistics. All 26 of the checklist items which covered content of the examinee's report were judged acceptable on the basis of

p-values and correlation with the other 25 items. A comparison between p-values obtained in a subsample of 59 Corps of Engineers Officers, who had test-relevant training and p-values for the remaining examinees did not reveal any items suspect in content or direction of scoring by reason of significantly lower scores for the engineer group.

Statistics of Major Variables and Effect of Order of Work. Major test statistics (means, standard deviations, and intercorrelatives) were notable chiefly for the strong correlation of order of work, that is, undertaking the site report first or runway first, with runway and total scores. These two scores tended to be lower if the examinee began with the site report. The negative relationships were attributed largely to two facts. 1) Most of those who began with the more time-consuming site report did not have time to complete the runway computation, or even begin it. 2) The scoring systems employed gave identical credit for application of an evaluative consideration regardless of whether it was applied to one, two, or all three sites. Therefore, time spent on a second and third site, though equally part of the assignment, was less productive of scoring credit than time spent on the runway. Rescoring the site reports was not undertaken because extensive time would be required. However, for approximate correction of the site-report score, a fraction of the obtained score proportional to the number of sites covered in the report was added to the score.

Effect of Examiner, Hour of Administration, and Calendar Period. Other potential sources of error--the examiner who conducted the test, hour of administration (3 A.M. vs. 7 P.M.), and calendar period (earlier vs. later months of testing)--were studied by analysis of variance. The site report score varied significantly with all three of these variables, scores averaging higher in the evening and in the earlier time period. The total score was significantly related only to calendar period. Correction of scores did not appear necessary, because magnitude of effects was not large. Moreover, the calendar-period effect could have been due to systematic change in examinees sampled rather than change in instruction or scoring standards.

Departures from Instructions. Some examinees, apparently having misunderstood instructions, prepared a report supporting one site rather than analyzing all, or presented judgments without the required factual basis. The multiple correlation coefficient of the report score with these two tendencies was -.41. The reduction in score represents in unknown proportion inadequate instruction of the examinee (which it would be desirable to correct), lack of examinee ability to comprehend and follow instructions (which might appropriately affect scoring), and deliberate shortcuts undertaken under time pressure.

Reliability. Only internal consistency measures of reliability were available. Such measures are inappropriate for a speeded test in which items are performed in a uniform sequence and, therefore, for the runway computation. However, time pressure was judged unlikely to have a systematic inflationary effect on internal consistency measures for the site report. The Kuder-Richardson 20 coefficient for the site report is .65. Under assumption that runway and site items are of similar reliability, the corresponding coefficient for the runway would be .54 (as compared with a directly obtained Kuder-Richardson of .84). The relatively low level of internal consistency reliability may be attributed to the relatively small number of scoring points together with some heterogeneity among them. The obtained measures of internal consistency do not reflect impairment in comparability of scores arising from sources of error, such as those discussed above, which affect the examinee's entire site, runway, or total score. The appropriateness of estimates of .65 and .54 was confirmed by somewhat lower Kuder-Richardson coefficients, affected by restriction of range, for the subsample that had completed both parts of the test.

Effect of Prior Experience. The extent to which prior experience may affect scores was investigated with respect to Army branch. In ten Army branches, the aggregate effect of branch was not great. The variance of branch means on site, runway, and total score was approximately 9%, 4%, and 8%, respectively, of individual-score variances. However, membership in the Corps of Engineers was associated with high scores, especially on runway (the engineer mean was a standard deviation above that of other examinees) and total scores (two-thirds a standard deviation above). The test thus appears sensitive to closely relevant experience. However, Corps of Engineers superiority may have arisen partly from higher test aptitude among those drawn into this branch.

Factor Analysis. A principal-component factor analysis of scores on the 26 checklist items made by the 294 examinees who finished both parts of the test, and varimax rotation, resulted in acceptance of a six-factor solution, which accounted for 49% of the total variance. Factors 2 (altitude and grade correction) and 4 (correct safety factor and consideration of surface softness) represent runway computation. The four factors describing site-report content were defined as follows: No. 1., operational hazards; No. 3., engineering and tactical considerations; No. 5., use or removal of buildings and vegetation; and No. 6., geographic considerations affecting site suitability. The higher loading items contributing to the definition and serving as a factor score were listed for each factor.

Total Score. For use in other analyses, a total score was formulated consisting of the site-report score, corrected for number of sites completed, and the runway score. The latter, which was most affected by the order in which the two parts of the task were performed, was given somewhat less relative weight than would be obtained by simple addition of raw scores.

Summary Evaluation. Scores on the Airfield Layout test may be strongly affected by random and systematic error, as indicated by fairly low internal consistency, demonstrated effects of certain circumstances of administration, and pronounced effect of the order in which the examinee undertook task requirements. However, basic scoring points appear fairly comprehensive and sound. In the test sample of several hundred, statistically significant correlation should be possible, despite attenuation, with any variables having substantial representation of abilities required in reporting on airfield sites or determining needed length of runway.