DEVELOPMENT AND CROSS-VALIDATION OF SCORING KEYS FOR LEADERS'
COURSE SELECTION INSTRUMENTS

PRS REPORT NO. 814

Foreword: PRS reports are primarily technical. While
conclusions affecting military policy or operations may
appear in them, they are not intended as a basis for
official action. Findings and conclusions contained in
PRS reports are intended to guide the conduct of further
research. When research findings suggest recommendations
for administrative action, such recommendations are made
separately to the appropriate military agency.

20 December 1949

Personnel Research Section
Personnel Research and Procedures Branch, AGO

BRIEF

STATEMENT OF PROBLEM:

Two of the instruments developed for selection of men for Leaders' Course were (1) a revision of the Biographical Information Blank for Officer Candidates (OCB), and (2) the Enlisted Man's Evaluation Report (LPE). This study is concerned with the further improvement of both of these instruments as predictors of success or failure in Leaders' Course.

RESULTS:

1. The OCB and the LPE in combination were more effective in differentiating between probable successes and failures in Leaders' Course $(r = .40)$ than either test used alone. Of the two, the LPE demonstrated greater validity for this purpose $(r = .35$ for LPE; $r = .25$ for OCB$)$.

2. A simplification in scoring the OCB was achieved by deleting fourteen sports-participation items with practically no reduction in validity (an increase from .29 to .31 with one experimental group used in an earlier analysis, a decrease from .41 to .40 with another).

3. An alternative method of item selection for the OCB, based on validity of the item against a composite rating on leadership, gave no appreciable increase in validity over the key in current use, the items of which had shown a consistency of validity in three different groups.

CONCLUSIONS:

1. Although the LPE demonstrated higher validity than the OCB, the validity of the instruments was higher when they were used in combination.

2. A simplification in scoring the OCB had no appreciable effect on the instrument's validity.

3. The alternative basis of selecting items for the OCB revision did not prove to be profitable.

WORK SUMMARY:

Scores that candidates had achieved in the OCB and LPE were compared with their final Leaders' Course standing. These data were used in (1) selecting the most valid items for both selection instruments, (2) determining the validities of these revised forms, both separately and in combination, and (3) deciding upon the best scoring methods.

DEVELOPMENT AND CROSS-VALIDATION OF SCORING KEYS FOR LEADERS' COURSE SELECTION INSTRUMENTS

BACKGROUND

Selection for Leaders Course is based in part on two instruments: (1) the Officer Candidate Biographical Information Blank (OCB-2 and OCB-3)[1]/ which is filled out by the candidate himself, and (2) the Enlisted Man's Evaluation Report, LPE-1, (DA AGO PRT-739), which is filled out by the enlisted man's superiors during his basic training.

Item analysis of the Officer Candidate Biographical Information Blank and the development of a scoring key for optimal selection of Leaders' Course applicants is described in PRS Report 764. As indicated in that report, the key which was developed showed validities of .29 and .41 against success in Leaders' Courses in two cross-validation populations. In subsequent preparation of a short form of the OCB, including only those items involved in the keyed responses, it was found that scoring could be reduced to a single run through a test-scoring machine involving only one side of an answer sheet, if one section of the OCB was omitted. This one section included keyed responses very similar to those which had been keyed in other sections of the OCB. Consequently, analysis was undertaken to determine whether deletion of these keyed responses would seriously affect the validity. Further item analyses whose objectives were primarily methodological in nature were also undertaken in connection with the OCB.

At the same time, item analysis of the Enlisted Man's Evaluation Report, LPE-1 was undertaken and a new scoring key was developed and cross-validated. This project (PJ 4105-11)[2]/ is primarily concerned with item analysis, and development and cross-validation of scoring keys for LPE-1. In order to provide a complete summary of all validation work accomplished on Leaders' Course selection instruments since preparation of PRS Report 764, the additional analysis of OCB-2 will be included in the present report.

The general objectives of this report, then, are: (1) to summarize further validation work on keys for the OCB, (2) to summarize the item analysis and development of a scoring key for LPE-1, and (3) to report results obtained in cross-validation of the scoring key for LPE-1, further cross-validation of the key for the OCB, and research to determine the optimal composite measure for Leaders' Course selection.

1/ Form OCB-2 (DA AGO PRT-648) had been employed in place of OCB-3 (DA AGO PRT-735) for Leaders' Course selection until available supplies of OCB-2 were exhausted. OCB-2 and -3 are nearly identical and, for the purposes of this paper, can be considered identical. Throughout this paper the simpler designation OCB will be employed.

2/ Disposition Form, File No. WDGPA 352, from D/P and A to TAG, Subject: "Potential Leaders' School(s)," dated 7 April 1947.

It should be pointed out at this time that all predictor and criterion data involved in the results here reported were obtained from the operating selection and training program in the various Leaders' Courses rather than from instruments administered for research purposes only.

METHOD

The populations for this study include the two cross-validation populations for OCB (see PRS Report 764), a population of 1,000 cases used for item analysis of LPE-1, and two populations (one of 500 cases and one of 170 cases) used for cross-validation of LPE-1 and OCB and for determination of the relative weighting of these two instruments in obtaining a composite score. LPE-1 papers were not available for the two cross-validation populations for OCB; OCB papers were not available for the 1,000 cases used for item analysis of LPE-1.

All cases are Leaders' Course graduates who had taken the selection instruments as applicants in the usual manner prior to entering the course. Some curtailment has occurred both in the distribution of scores on the selection instruments (since not all applicants are accepted) and on criterion variables (because of failure in the Leaders' Course). The effect of such restriction although it cannot be estimated exactly, would be a reduction in the degree of relationship obtained between the predictor and criterion variables described below.

Predictor Variables

1. Biographical Information Blank, OCB-2 (DA AGO PRT-648) or OCB-3 (DA AGO PRT-735) consists of 91 questions on background and past experience (Part I), 28 preference check list pairs (Part II), 52 preference check list most-least quintets (Part III), and 24 preference check list pairs (Part IV). In part IV the subject not only indicates his choice between the two members of a forced choice pair but also indicates whether only the member of the pair that he marked applies to him, or both members apply to him, or neither member applies to him.

2. The Enlisted Man's Evaluation Report, LPE-1 (WD AGO PRT-739) is filled out by the training division NCO who is designated as best able to rate the trainee. The form is indorsed by the platoon leader or company commander. It consists of the following sections:

Section I. Instructions to Adjutants or personnel officers administering the rating form.

Section II. Twenty-five groups of preference check list tetrads completed by rater only. The "Most Descriptive" and the "Least Descriptive" member of each tetrad is indicated by the rater.

Section III. A 20-point rating scale on over-all competence as a prospective platoon sergeant, accomplished by rater.

Section IV. Two 5-point rating scales, one on demonstrated leadership and one on character, accomplished by rater.

- 6 -

Section V. Period covered by report, summary of chief assignments during the rating period, indication of the degree to which the rater believes he is qualified to rate the ratee, comments, and authorization -- all to be entered by the rater.

Section VI. A 20-point scale on over-all competence as a prospective platoon sergeant. This section corresponds to Section III, but is entered by the indorser. The indorser is instructed that, if he concurs with the entry in Section III by the rater, he will make the same entry as that of the rater; if he disagrees, he will make the entry he believes appropriate.

Section VII. Two 5-point scales on demonstrated leadership and character, accomplished by the indorser. These correspond in content to the scales of Section IV and the directions are the same as for Section VI.

Criterion Variables

The official Leaders' Course grade (final score at Leaders' Course) was used as the criterion for the analyses to be described. The component measures are:

1. A rating of proficiency as acting NCO in Phase II.[3] This rating is made by the commander of the company to which the man is assigned, using the Leaders' Course Board Rating and Report Form (DA AGO PRT-1621). This form includes 20 most-least forced choice list items and 20 5-point rating scale items, together with an over-all rating.

2. Ratings on leadership characteristics made by officer instructors of the Leaders' School which employ the form DA AGO PRT-1621. These ratings are concerned with performance during Phase I.[3]

3. A rating made by fellow students in Phase I on the Student Leadership Evaluation Report, (DA AGO PRT-829). These ratings are made by groups of 8 to 15 men who have good opportunity to know each other over a three weeks' period.

4. The Leaders Reaction Test, a field-type situational test in which subjects are rated by trained observers on their leadership behavior during a series of specified situations.

Effect of the Deletion of Items 78-91 (OCB, Part I) on the Validity of the Leaders' Course Item Analysis Key

Items 78 through 91 are a series of questions requiring ratings on degree of interest and participation in various sports. Elsewhere in Part I, background items covered this type of contest with reasonable

---

[3] The Leaders' Courses are divided into two phases. During Phase I, students receive instruction in the principles of leadership at the Leaders' Schools. During Phase II, they are assigned to training companies and given opportunity to apply the principles they have learned during Phase I. Each phase is normally of three weeks' duration.

thoroughness. These items required a complete set of directions and considerable space on the scoring sheet. It was decided to rescore the OCB papers of the cross-validation samples (see PRS Report 764) in order to determine the effect of the deletion of these items on the validity of the key. The following two samples were used: (1) 180 Leaders' Course graduates from Fort Jackson and, (2) 146 Leaders' Course graduates from Fort Ord. Validities of the complete item analysis key and of the key after deletion of the fourteen items in question are given in Table 1 below.

TABLE 1

VALIDITY OF THE OCB ITEM ANALYSIS KEY AGAINST FINAL SCORE IN TWO
POPULATIONS OF LEADERS COURSE GRADUATES BEFORE AND AFTER
DELETION OF FOURTEEN SPORTS PARTICIPATION ITEMS

	Ft. Jackson N = 180	Ft. Ord N = 146
Before Deletion	.41	.29
After Deletion	.40	.31

Deletion of these fourteen items has a negligible effect on the validity of the key. In the Fort Jackson population, a decrease of .01 was obtained. In the Fort Ord population, an increase of .02 was obtained. The effect, if any, was to increase the validity. Consequently, it was decided to exclude the items in revision of the OCB.

A Comparison of Two OCB Keys Developed by Two Different Item Selection
Methods

A second study was undertaken partly for methodological purposes and partly in the hope of improving upon the validity of the item analysis key by rekeying according to a different item selection method. Biserial correlations against an external criterion (three ratings on leadership) were the indexes of item validity in the first sample. The difference in percentage of high rated men and low rated men selecting each item alternative divided by the number of cases ($\frac{H-L}{N}$) was used in the second and third samples. A complete description of the first procedure employed in selecting items for keying may be obtained from PRS Report 764. In brief, items were selected according to the consistency of their validity in three item analysis populations. This method of item selection appears, from a theoretical viewpoint, to have disadvantages similar to those involved in the requirement that an applicant obtain a given score on each selection instrument (multiple cutting scores). If all items having a validity of .10 or greater in each of three item analysis samples are selected, it would be possible to accept an item having validities of .10, .10, and .10 while rejecting a second item having validities of .30, .40, and .09, although the second of these two items is obviously superior.

Because of the belief that these apparent difficulties in the procedure of the first item analysis could be alleviated and in order to obtain empirical data to aid in deciding whether this consistency method of selecting items is actually inferior to selection on the average validity, the item validity indices in the two item analysis populations employing the high minus low index were recomputed as tetrachoric correlations. A single average validity coefficient was then obtained employing the two tetrachoric coefficients and the biserial coefficient already available in the third item analysis sample. A new key was then developed on the basis of the average validity index. The two cross-validation populations which were utilized for the analysis leading to the deletion of the 14 sports participation items were employed again for this additional analysis. This new key is referred to as the Averages Key. In Table 2, validities of the Consistency Key are compared with those of the Averages Key. Additional coefficients, explained in the following paragraph, were also included.

In PRS Report 764, it was shown that a combination of the Consistency Key and the Current Operating Key[4]/ (developed for officer candidate school selection) gave higher validity for the OCB than either alone, and it was proposed that the two keys be combined for future operating use. Because of this finding, it seemed pertinent to determine the validity of the Averages Key in combination with the Current Operating Key in order that the resulting validities could be compared with those obtained in combining the Consistency Key and the Current Operating Key. The Current Operating Key has been used for Leaders' Course selection pending the results of the present study. The required correlations of sums were computed and are reported in Table 2.

TABLE 2

COMPARISON OF VALIDITIES OF AN ITEM ANALYSIS KEY DEVELOPED BY
SELECTING ITEMS CONSISTENTLY VALID IN THREE POPULATIONS
WITH VALIDITIES OF A KEY DEVELOPED ON THE BASIS OF
AVERAGE VALIDITY IN THREE POPULATIONS

	Ft. Jackson N = 180		Ft. Ord N = 146	
	Averages Key	Consistency Key	Averages Key	Consistency Key
Rights	.37	.34	.18	.27
Wrongs	-.28	-.14	-.21	-.21
R - W	.38	.34	.23	.29
Current Operating Key	.41	.41	.29	.29

The coefficients reported in Table 2 suggest that the method of item selection has little or no effect on the cross-validated validity. The

4/ PRS Report 752, Procedures for Selection of Enlisted Men for Officer Training, The Adjutant General's Office, Personnel Research Section, 9 February 1948.

absence of improved validity with the Averages Key led to the decision that the Consistency Key (in combination with the Current Operating Key) would be recommended for operational use.

Development of Scoring Procedures for the Enlisted Man's Evaluation Report, LPE-1, (DA AGO PRT-739)

It will be noted in the description of LPE-1 (pp. 6-7) that this rating form includes a set of forced choice items, a set of two 5-point scales, and a 20-point scale, all to be entered by the rater; and a set of two 5-point scales and a 20-point scale to be entered by the indorser. As in the case of the Current Operating Key for OCB, the Current Operating Key for the forced choice section of LPE-1 was originally developed for OCS selection. The item validity indices were computed against an associates' rating criterion among recruits who were potential officer candidates but not actually in officer candidate school.

The first step in evaluating and improving the key and method of scoring for LPE-1 was the determination of the validity of the several sections of the rating form by present scoring procedures. This analysis had three objectives, namely: (1) development of the "best" composite of the graphic scales, (2) evaluation of the adequacy of the Current Operating Key for forced choice items and, (3) evaluation of the expectancy of validity for LPE-1 as a whole.

The variables of this correlational analysis are:

1. Final Score at Leaders' Course -- the criterion.

2. Section II, (forced choice) by rater -- Current Operating Key.

3. Section III. A 20-point scale on over-all competence as a prospective platoon sergeant, entered by rater.

4. Section IV. A 5-point scale on demonstrated leadership (Scale A), entered by rater.

5. Section IV. A 5-point scale on character (Scale B), entered by rater.

6. Section VI. A 20-point scale on over-all competence as a prospective platoon sergeant, entered by indorser.

7. Section VII. A 5-point scale on demonstrated leadership (Scale A), entered by indorser.

8. Section VII. A 5-point scale on character (Scale B), entered by indorser.

A population consisting of 1,000 Leaders' Course graduates from Ft. Ord, Ft. Knox, Ft. Dix, Ft. Jackson, and Camp Breckenridge was employed for this analysis. LPE-1 had been administered to this population when the men were applicants for entry to the Leaders' Course during a period extending from February 1948 to September 1948. The means, standard deviations, and intercorrelations of these variables are presented in Table 3.

TABLE 3

MEANS, STANDARD DEVIATIONS, INTERCORRELATIONS, AND VALIDITIES OF LPE-1
PART SCORES FOR 1,000 LEADERS' COURSE GRADUATES

Variable	Mean	Standard Deviation	Description of Variables	Intercorrelations
1.	70.9	6.3	Final Score, Leaders' Course	1
2.	3.1	9.7	Sec. II, (Rater) Preference Check List	.09 2
3.	12.5	4.1	Sec. III, (Rater) Over-all Competence	.18 .30 3
4.	3.2	.9	Sec. IV, A (Rater) Demonstrated Leadership	.18 .30 .64 4
5.	4.1	.9	Sec. IV, B (Rater) Character	.07 .35 .39 .47 5
6.	12.4	3.9	Sec. VI, (Indorser) Over-all Competence	.20 .29 .90 .62 .40 6
7.	3.2	.9	Sec. VIII, (Indorser) Demonstrated Leadership	.16 .30 .61 .87 .44 .67 7
8.	4.1	.9	Sec. VII, (Indorser) Character	.08 .32 .34 .42 .78 .39 .44

Several observations can be made from an examination of Table 3. First, the Current Operating Key for the forced choice items (Section II) is inadequate (validity coefficient of .09). Secondly, regarding the relative usefulness of the several graphic scales, Scale B (Character), whether filled out by rater (Section IV, B) or by indorser (Section VII), has lower validity than the remaining graphic scales (Over-all Competence and Demonstrated Leadership). Sections III and VI (Over-all Competence) by rater and indorser appear to have somewhat higher validity than Sections IV and VII (Demonstrated Leadership). In addition, the correlation between Sections III and IV (.64) and between Sections VI and VII (.67) is sufficiently high that Sections IV and VII add little or nothing to the validity of the composite. It was decided, consequently, that a combination of Sections III and VI would give an optimal composite.

Biserial validities of the 200 alternatives (8 per item) of the forced choice items of Section II were computed for the population of 1,000 cases from Ft. Ord, Ft. Knox, Ft. Dix, Ft. Jackson, and Camp Breckenridge.

Those items having highest validity were selected for the item analysis key. Because of the greater presumptive unreliability of item biserial validities having P values (percentage of raters selecting an item alternative) less than .20 or greater than .80, more stringent requirements were set for the inclusion

of such items in the final key. An additional modification was made because
it was found that the mean item validities for "Most Descriptive" alternatives
were positive while those for "Least Descriptive" alternatives were negative.
The mean validity of a group of items selected as most valid in an item anal-
ysis sample will regress toward their mean value in a cross-validation
population.[5] The cross-validated validity is, of course, in large part
determined by the mean validity of the keyed items in the cross-validation
population. Consequently, allowances were made for the expected effect of
this regression phenomenon on the mean validities in the cross-validation
population of the alternatives included in the rights key for the "most"
alternatives, the wrongs key for th) "most" alternatives, the rights key for
the "least" alternatives, and the wrongs key for the "least" alternatives.
The allowance was intended to yield equal mean regressed validities for each
grouping.

Taking into account allowance for the greater unreliability of item
validities at extreme points of cut ($>$ .80 or $<$ .20) and the regression
effects just considered, the following standards for item selection were set:

TABLE 4

STANDARDS FOR ITEM SELECTION

Alternative	P Values		Key	Requisite Validity
Most	.20 to	.80	Rights	.08 to 1.00
Most	$>$.80 or	$<$.20	Rights	.13 to 1.00
Most	.20 to	.80	Wrongs	-.13 to 1.00
Most	$>$.80 or	$<$.20	Wrongs	-.18 to -1.00
Least	.80 to	.20	Rights	.10 to 1.00
Least	$>$.80 or	$<$.20	Rights	.18 to 1.00
Least	.80 to	.20	Wrongs	-.14 to -1.00
Least	$>$.80 or	$<$.20	Wrongs	-.10 to -1.00

[5] The extent of the regression is determined by the slope of the regression line
and by the mean value. If the mean value of a total pool of items were zero,
an equal amount of regression would be expected with a selected group of items
having an average validity of +.30 and a second group having an average validi-
ty of -.30. With the mean of the total pool equalling .10 and the slope of
the regression line .5, the regressed mean value of a group in the analysis
sample would be .20, while for a group with a mean value of -.30 in the item
analysis sample, the corresponding regressed mean value would be -.10. It
follows that, in the example cited, those items with a mean of +.30 would
on cross-validation contribute more than those with a mean of -.30.

Cross-Validation of OCB and LPE-1 Keys and Determination of an Optimal Composite for Selection Purposes

To obtain cross-validation of the item analysis key for Section II of LPE-1, to obtain further cross-validation of the Consistency Key for OCB, and to determine proper weighting of these two keys and the graphic scales of LPE-1 (Sections III and VI), an additional population of 500 cases was isolated. This population consisted of Leaders' Course graduates from Ft. Knox, Ft. Jackson, Ft. Ord, and Camp Chaffee. OCB and LPE-1 forms were accomplished on these men during a period extending from August 1948 to January 1949.

Criterion scores, OCB scores, and two scores on LPE-1 (Section II and the sum of scores on Sections III and IV) were determined for each individual. In processing the LPE-1 papers, it was discovered that the period of observation available to the raters (length of time LPE-1 rater was associated with the applicant at basic training) varied considerably. In a large portion of the cases this period of observation appeared too small to allow the rater to become reasonably well acquainted with the Leaders' Course applicant. This situation arose because the time allotted to basic training was reduced from 13 to 8 weeks. Since the superior officer of the recruit who applied for entry to Leaders' Course accomplished LPE-1 at least two weeks prior to recruit's completion of basic training, the period of observation allowed often was quite short. Because of this finding, it was decided to subdivide the 500 cases into those whose LPE-1 ratings were based upon an observation period of 52 days or more, those whose LPE-1 ratings were based on observation periods ranging from 23 through 51 days, and those whose LPE-1 ratings were based on periods of observation of 22 days or less.

TABLE 5

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF OCB (CONSISTENCY KEY), LPE-1 (SECTION II - FORCED CHOICE - ITEM ANALYSIS KEY), LPE-1 (SECTION III PLUS SECTION VI) AND FINAL SCORE AT LEADERS' COURSES FOR A CROSS-VALIDATION POPULATION OF 500 LEADERS' COURSE GRADUATES SUBDIVIDED ACCORDING TO LENGTH OF OBSERVATION PERIOD ON LPE-1

Observation Period*	Description of Variables	Mean	Standard Deviation	Intercorrelations of				
				1. Final Score	2. LPE-1 (II-FCL Key)	3. LPE-1 (III VI)	4. LPE-1 Total Composite	5. OCB (Consistency Key)
1 - 22 days N = 118	1. Final Score	74.8	5.7	---				
	2. LPE-1, (II-FCL Key)	2.2	4.0	.18	---			
	3. LPE-1, (III+VI)	25.3	7.2	.14	.40	---		
	4. LPE-1, (Composite)	14.8	6.3	.19	**	**	---	
	5. OCB (Consistency Key)	41.8	9.0	.27	.18	.18	.22	---
23 - 51 days N = 334	1. Final Score	72.9	6.2	---				
	2. LPE-1, (II-FCL Key)	3.0	4.3	.08	---			
	3. LPE-1, (III+VI)	26.0	7.5	.09	.40	---		
	4. LPE-1, (Composite)	16.0	6.7	.10	**	**	---	
	5. OCB (Consistency Key)	42.5	8.6	.22	-.04	-.07	-.06	---
52 days N = 44	1. Final Score	69.1	6.3	---				
	2. LPE-1, (II-FCL Key)	1.4	4.5	.42	---			
	3. LPE-1, (III+VI)	24.2	7.0	.42	.72	---		
	4. LPE-1, (Composite)	13.5	7.4	.45	**	**	---	
	5. OCB (Consistency Key)	42.6	9.1	.38	.24	.10	.19	---

* Length of time LPE-1 rater was associated with applicant at basic training.

** Part-whole correlations, not computed.

- 14 -

Within these three subgroupings, intercorrelations of OCB (Consistency Key), LPE-1 (Sect. II-item analysis key), LPE-1 (Sect. III + Sect. VI) and the Final Score in Leaders' Course were computed. These are presented in Table 5 together with the validity of a composite score for LPE-1 $/.5$ (III + VI) + II$/$ and its correlation with OCB (Consistency Key). The raw score weights of unity and .5 for II (forced choice items) and III + VI (graphic scales) respectively, give equal standard score weighting of these two components. It will be noted in Table 5 that the validities of the two components are approximately equal while the standard deviation of the graphic scales is approximately double that of the forced choice section.

The most striking finding in Table 5 is the low validity of the LPE-1 scores in the two shorter periods of observation. While this finding is in line with the expected effect of the shorter period of observation, it does not seem to be explainable solely on this basis. First of all, the validities are lower in the 23-51 days group than in the 1-22 days group. Secondly, OCB has lower validity in these two groups than in the third grouping and lower validities than had been obtained with prior cross-validation populations.

The negative correlations between OCB and the LPE-1 scores in the 23-51 days group is an unusual finding which may have bearing upon the low LPE-1 and OCB validities. If restriction in range had been imposed by selecting on a composite of LPE-1 and OCB, negative correlations within the restricted group would be expected. Of course, restriction could not have occurred directly on a composite of the LPE and OCB scores as defined in Table 5 since, in the operating selection program, the scoring keys were different. However, the OCB and LPE scores of Table 5 can be expected to show high correlation with those used in the operating selection program; because of this high correlation, selection on the composite used in the operating program would tend to produce the negative correlations mentioned above. While this is a possible explanation of the low values obtained, the problem raised cannot be adequately solved without further analysis on a new cross-validation population.

This considerable drop in the validities for the shorter period does suggest that care should be taken to insure adequate observation periods in obtaining any type of rating evaluation, but this conclusion cannot be regarded as well established empirically.

Proper weighting of the LPE and OCB experimental keys to obtain a composite with optimal validity presents something of a problem. The data of Table 5, if the LPE-1 validities for all three periods of observation are considered, probably show bias favoring OCB in comparing the relative validity of the two instruments. If consideration is limited to the cases with adequate observation period, the N of the sample involved is too small to allow reliable determination of their proper weighting. Equal raw score weighting was chosen as being administratively most simple, and, on the basis of general subjective judgment, as giving the most nearly optimal combination.

The means, standard deviations, and validities of the equally weighted composite is given in Table 6 for the three sub-groups of Table 5.

TABLE 6

MEANS, STANDARD DEVIATIONS, AND VALIDITY OF AN EQUALLY WEIGHTED
COMPOSITE OF OCB AND LPE-1 EXPERIMENTAL KEYS AGAINST FINAL
SCORE IN LEADERS COURSE FOR 500 LEADERS' COURSE
GRADUATES SUBDIVIDED ACCORDING TO LENGTH OF
OBSERVATION PERIOD ON LPE-1

Observation Period	N	Mean	SD	Validity
1 - 22 days	118	56.6	12.1	.30
23 - 51 days	334	58.5	10.4	.25
52> days	44	56.1	12.8	.53

Further Cross-Validation of OCB[6]/ and LPE-1 Keys and of the Composite Score

Because of the inadequacies in the data of the first cross-validation population, a second population of 170 Leaders' Course graduates from Camp Chaffee, Ft. Ord, and Ft. Knox were isolated. All members of this second population had entered the Leaders' Course after the longer (14-week) basic training cycle had been reinstalled. The period of observation for the LPE-1 rater was adequate in all instances.

Total scores were determined for both the LPE-1 and the OCB using the experimental keys, and the intercorrelations among these total scores and the Final Score at the Leaders Course were computed. These intercorrelations are presented in Table 7.

TABLE 7

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF THE EXPERIMENTAL
KEYS FOR OCB, THE EXPERIMENTAL KEY FOR LPE-1, AND FINAL SCORE AT
LEADERS' COURSE FOR 170 LEADERS' COURSE GRADUATES FROM
CAMP CHAFFEE, FT. KNOX, AND FT. ORD

Variable	Mean	SD	Intercorrelations 1. OCB	Score 2. LPE-1
1. OCB	43.3	7.1	1	
2. LPE-1	16.1	7.5	.12	2
3. Final Score	69.7	6.2	.25	.35
Validity of OCB + LPE-1 = .40				

6/ The key for OCB has been revised since the first cross-validation analysis described for reasons of administrative feasibility. The revision consisted in the deletion of a section of items relating to prior service and a section of items relating to university training both of which were considered to be not usually pertinent in the present population of applicants. The effect of this revision on the validity of the key is not known.

While the validity of OCB against Final Score (.25) as shown in Table 7 is disappointingly low, the validity of LPE-1 against Final Score (.35) is adequate but not outstanding. Because of the low correlation between the scores on the two instruments, the validity of the sum of the two instruments is appreciably higher than the validity of either one alone.

Since the validity of LPE-1 is higher than that for OCB, it is evident that with heavier weighting of the former, the sum of the two instruments would show somewhat higher validity. However, with the optimal weights provided by mulitple correlation, the validity of the composite is .41, a negligible increase of .01 over that provided by equal weighting of both instruments. Since consideration of all available validity information on both instruments leads to no great confidence in the higher validity of LPE-1, it appeared desirable to retain the administratively more convenient equal raw score weighting.

In discussing the first of the two cross-validation studies (p. 6), the probable effect of restriction of range of the predictor instruments on the obtained validity coefficients was briefly considered. The effect of such restriction in range on validity was more carefully examined in this second cross-validation study. It was found that restriction on the predictor instruments OCB and LPE-1 could not be determined in a satisfactorily accurate manner. The difficulty arose principally from the long delay which frequently occurred between testing of the applicants and admission to the school. Rejection of an applicant could not be assumed because he failed to appear in a class beginning shortly after administration of the instruments.

It was discovered, however, that the degree of restriction on the criterion (Final Score) was considerable -- about 30% of the population of the study in question failed to graduate. To provide a rough estimate of the effect of such restriction, the validity of the composite measure was corrected for restriction in range using the formula

$$r_{corr} = \frac{rk}{\sqrt{1 - (1-k^2)\ r^2}}$$

where k is the ratio of the standard deviation of the variable on which restriction occurred in the unrestricted sample to its SD in the restricted sample. Since data for computing the SD in the unrestricted sample were not available, this ratio was computed by assuming normal distribution of the criterion.[7]

The corrected coefficient was .62. While it is not believed desirable to employ this coefficient as an estimate of the "true" validity of the composite score in the unrestricted sample, the increase of .22 obtained will serve to emphasize the fact that restriction does seriously reduce the estimates of validities obtained from validation analysis within a population of Leaders' Course graduates.

7/ Taylor, E. K. and Gaylord, R. H. Table for use in the computation of statistics of dichotomous and truncated distributions. Educ. Psychol. Meas., 1947, 7, pp 441-456.

This estimate is not regarded as accurate for two reasons: (1) it is doubtful if the assumption involved in the use of the formula -- that individuals who failed to graduate were low on the criterion variables -- is legitimate. Separation of students often occurs early in the course and is not based on all the evaluative measures used in determining Final Score, (2) rejection on the selection instruments is a source of restriction not considered in this connection.

CONCLUSIONS

A. Deletion of 14 sports participation items in Part I of the Biographical Information Blank for officer candidates (OCB) did not appreciably affect the validity of the total key.

B. Rekeying the OCB with selection of items according to their average validity in three item analysis populations resulted in no improvement in validity over that obtained with a key developed with selection of items according to consistency of validity among the three populations.

C. Further cross-validation of the experimental key for the OCB gave validities against Final Score at Leaders' Course for four groupings of .27, .22, .38, (Table 5) and .25, (Table 7). When these validities are weighted according to the number of cases in each grouping (118, 334, 44, and 170), the resulting validity was .24. This is substantially lower than the values of .40 and .31 (Table 1) obtained in prior cross-validation. An average of all cross-validation groups gives .28 as the best over-all estimate of the validity of the OCB experimental key.

D. Cross-validated validities against Final Score of the new key for LPE-1 of .45 (52 days, Table 5) and .35 (Table 7) were obtained for the groups in which the rater had adequate observation.

E. Length of the period of observation available to the rater appears to have an appreciable effect on the validity of his evaluations.

F. Improvement over the validity obtained with the current operating keys appears well established in the case of both OCB and LPE.

Personnel

A. Program Coordinator - H. E. Brogden

B. Statistical Advisor - Bertha Harper

C. Project Director - H. E. Brogden

D. Preparation of Report - H. E. Brogden